(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
14 February 2002 (14.02.2002)

PCT

(10) International Publication Number
WO 02/12340 A2

(51) International Patent Classification[7]: C07K 14/705

(21) International Application Number: PCT/US01/24217

(22) International Filing Date: 1 August 2001 (01.08.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

60/223,269	3 August 2000 (03.08.2000)	US
60/224,456	10 August 2000 (10.08.2000)	US
60/226,410	18 August 2000 (18.08.2000)	US
60/228,140	25 August 2000 (25.08.2000)	US
60/230,067	31 August 2000 (31.08.2000)	US
60/231,434	8 September 2000 (08.09.2000)	US

(71) Applicant *(for all designated States except US)*: **INCYTE GENOMICS, INC.** [US/US]; 3160 Porter Drive, Palo Alto, CA 94304 (US).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: **YUE, Henry** [US/US]; 826 Lois Avenue, Sunnyvale, CA 94087 (US). **THORNTON, Michael** [US/US]; 9 Medway Road, Woodside, CA 94062 (US). **RAMKUMAR, Jayalaxmi** [IN/US]; 34359 Maybird Circle, Fremont, CA 94555 (US). **TANG, Y., Tom** [US/US]; 4230 Ranwick Court, San Jose, CA 95118 (US). **AZIMZAI, Yalda** [US/US]; 5518 Boulder Canyon Drive, Castro Valley, CA 94522 (US). **BAUGHN, Mariah, R.** [US/US]; 14244 Santiago Road, San Leandro, CA 94577 (US). **YANG, Junming** [CN/US]; 7125 Bank Lane, San Jose, CA 95129 (US). **YAO, Monique, G.** [US/US]; 111 Frederick Court, Mountain View, CA 94043 (US). **LAL, Preeti** [IN/US]; P.O. Box 5142, Santa Clara, CA 95056 (US). **WALIA, Narinder, K.** [US/US]; 890 Roble Avenue #1, San Leandro, CA 94577 (US). **GANDHI, Ameena, R.** [US/US]; 837 Roble Avenue, #1, Menlo Park, CA 94025 (US). **HAFALIA, April, J., A.** [US/US]; 2227 Calle de Primavera, Santa Clara, CA 95054 (US). **NGUYEN, Danniel, B.** [US/US]; 1403 Ridgewood Drive, San Jose, CA 95118 (US). **PATTERSON, Chandrea** [US/US]; 490 Sherwood Way #1, Menlo Park, CA 94025 (US). **ELLIOTT, Vicki, S.** [US/US]; 3770 Polten Place Way, San Jose, CA 95121 (US). **TRIBOULEY, Catherine, M.** [FR/US]; 1121 Tennessee, #5, San Francisco, CA 94107 (US). **LU, DYUNG, Aina, M.** [US/US]; 233 Coy Drive, San Jose, CA 95123 (US). **XU, Yuming** [US/US]; 1739 Walnut Drive, Mountain View, CA 94040 (US). **REDDY, Roopa** [IN/US]; 1233 W. McKinely Avenue #3, Sunnyvale, CA 94086 (US). **HERNANDEZ, Roberto** [GB/GB]; 160 Pine Tree Avenue, Canterbury Kent, CT2 7SZ (GB). **BOROWSKY, Mark, L.** [US/US]; 122 Orchard Avenue, Redwood City, CA 94061 (US). **LO, Terence, P.** [CA/US]; 1451 Beach Park Boulevard, Apt. 115, Foster City, CA 94404 (US). **LU, Yan** [CN/US]; 3885 Corrina Way, Palo Alto, CA 94303 (US). **POLICKY, Jennifer, L.** [US/US]; 1511 Jarvis Court, San Jose, CA 95118 (US). **GREENE, Barrie, D.** [US/US]; 1332 10th Avenue #104, San Franciso, CA 94122 (US). **SANJANWALA, Madhu, S.** [US/US]; 210 Sylvia Court, Los Altos, CA 94024 (US). **RAUMANN, Brigitte, E.** [US/US]; 5801 South Dorchester Avenue #3B, Chicago, IL 60637 (US). **BURFORD, Neil** [GB/US]; 105 Wildwood Circle, Durham, CT 06422 (US). **ISON, Craig, H.** [US/US]; 1242 Weathersfiedl Way, San Jose, CA 95118 (US). **LEE, Ernestine, A.** [US/US]; 624 Kains Street, Albany, CA 94706 (US). **DING, Li** [CN/US]; 3353 Alma Street #146, Palo Alto, CA 94306 (US). **DAS, Debopriya** [IN/US]; 1179 Bonita Avenue Apt. 3, Mountain View, CA 94040 (US). **KALLICK, Deborah, A.** [US/US]; 900 Olive Street, Menlo Park, CA 94025 (US). **KHAN, Farrah, A.** [IN/US]; 3617 Central Road #102, Glenview, IL 60025 (US). **SEILHAMER, Jeffrey, J.** [US/US]; 12555 La Cresta, Los Altos, CA 94022 (US).

(74) Agents: **HAMLET-COX, Diana** et al.; Incyte Pharmaceuticals, Inc., 3160 Porter Drive, Palo Alto, CA 94304 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

[Continued on next page]

(54) Title: TRANSPORTERS AND ION CHANNELS

(57) Abstract: The invention provides human transporters and ion channels (TRICH) and polynucleotides which identify and encode TRICH. The invention also provides expression vectors, host cells, antibodies, agonists, and antagonists. The invention also provides methods for diagnosing, treating, or preventing disorders associated with aberrant expression of TRICH.

WO 02/12340 A2

Published:
— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

TRANSPORTERS AND ION CHANNELS

TECHNICAL FIELD

This invention relates to nucleic acid and amino acid sequences of transporters and ion channels and to the use of these sequences in the diagnosis, treatment, and prevention of transport, neurological, muscle, immunological, and cell proliferative disorders, and in the assessment of the effects of exogenous compounds on the expression of nucleic acid and amino acid sequences of transporters and ion channels.

BACKGROUND OF THE INVENTION

Eukaryotic cells are surrounded and subdivided into functionally distinct organelles by hydrophobic lipid bilayer membranes which are highly impermeable to most polar molecules. Cells and organelles require transport proteins to import and export essential nutrients and metal ions including K^+, NH_4^+, P_i, SO_4^{2-}, sugars, and vitamins, as well as various metabolic waste products. Transport proteins also play roles in antibiotic resistance, toxin secretion, ion balance, synaptic neurotransmission, kidney function, intestinal absorption, tumor growth, and other diverse cell functions (Griffith, J. and C. Sansom (1998) The Transporter Facts Book, Academic Press, San Diego CA, pp. 3-29). Transport can occur by a passive concentration-dependent mechanism, or can be linked to an energy source such as ATP hydrolysis or an ion gradient. Proteins that function in transport include carrier proteins, which bind to a specific solute and undergo a conformational change that translocates the bound solute across the membrane, and channel proteins, which form hydrophilic pores that allow specific solutes to diffuse through the membrane down an electrochemical solute gradient.

Carrier proteins which transport a single solute from one side of the membrane to the other are called uniporters. In contrast, coupled transporters link the transfer of one solute with simultaneous or sequential transfer of a second solute, either in the same direction (symport) or in the opposite direction (antiport). For example, intestinal and kidney epithelium contains a variety of symporter systems driven by the sodium gradient that exists across the plasma membrane. Sodium moves into the cell down its electrochemical gradient and brings the solute into the cell with it. The sodium gradient that provides the driving force for solute uptake is maintained by the ubiquitous Na^+/K^+ ATPase system. Sodium-coupled transporters include the mammalian glucose transporter (SGLT1), iodide transporter (NIS), and multivitamin transporter (SMVT). All three transporters have twelve putative transmembrane segments, extracellular glycosylation sites, and cytoplasmically-oriented N- and C-termini. NIS plays a crucial role in the evaluation, diagnosis, and treatment of various thyroid pathologies because it is the molecular basis for radioiodide thyroid-imaging

techniques and for specific targeting of radioisotopes to the thyroid gland (Levy, O. et al. (1997) Proc. Natl. Acad. Sci. USA 94:5568-5573). SMVT is expressed in the intestinal mucosa, kidney, and placenta, and is implicated in the transport of the water-soluble vitamins, e.g., biotin and pantothenate (Prasad, P.D. et al. (1998) J. Biol. Chem. 273:7501-7506).

One of the largest families of transporters is the major facilitator superfamily (MFS), also called the uniporter-symporter-antiporter family. MFS transporters are single polypeptide carriers that transport small solutes in response to ion gradients. Members of the MFS are found in all classes of living organisms, and include transporters for sugars, oligosaccharides, phosphates, nitrates, nucleosides, monocarboxylates, and drugs. MFS transporters found in eukaryotes all have a structure comprising 12 transmembrane segments (Pao, S.S. et al. (1998) Microbiol. Molec. Biol. Rev. 62:1-34). The largest family of MFS transporters is the sugar transporter family, which includes the seven glucose transporters (GLUT1-GLUT7) found in humans that are required for the transport of glucose and other hexose sugars. These glucose transport proteins have unique tissue distributions and physiological functions. GLUT1 provides many cell types with their basal glucose requirements and transports glucose across epithelial and endothelial barrier tissues; GLUT2 facilitates glucose uptake or efflux from the liver; GLUT3 regulates glucose supply to neurons; GLUT4 is responsible for insulin-regulated glucose disposal; and GLUT5 regulates fructose uptake into skeletal muscle. Defects in glucose transporters are involved in a recently identified neurological syndrome causing infantile seizures and developmental delay, as well as glycogen storage disease, Fanconi-Bickel syndrome, and non-insulin-dependent diabetes mellitus (Mueckler, M. (1994) Eur. J. Biochem. 219:713-725; Longo, N. and L.J. Elsas (1998) Adv. Pediatr. 45:293-313).

Monocarboxylate anion transporters are proton-coupled symporters with a broad substrate specificity that includes L-lactate, pyruvate, and the ketone bodies acetate, acetoacetate, and beta-hydroxybutyrate. At least seven isoforms have been identified to date. The isoforms are predicted to have twelve transmembrane (TM) helical domains with a large intracellular loop between TM6 and TM7, and play a critical role in maintaining intracellular pH by removing the protons that are produced stoichiometrically with lactate during glycolysis. The best characterized H^+-monocarboxylate transporter is that of the erythrocyte membrane, which transports L-lactate and a wide range of other aliphatic monocarboxylates. Other cells possess H^+-linked monocarboxylate transporters with differing substrate and inhibitor selectivities. In particular, cardiac muscle and tumor cells have transporters that differ in their K_m values for certain substrates, including stereoselectivity for L- over D-lactate, and in their sensitivity to inhibitors. There are Na^+-monocarboxylate cotransporters on the luminal surface of intestinal and kidney epithelia, which allow the uptake of lactate, pyruvate, and ketone bodies in these tissues. In addition, there are specific and selective transporters for organic cations and organic anions in organs including the

kidney, intestine and liver. Organic anion transporters are selective for hydrophobic, charged molecules with electron-attracting side groups. Organic cation transporters, such as the ammonium transporter, mediate the secretion of a variety of drugs and endogenous metabolites, and contribute to the maintenance of intercellular pH (Poole, R.C. and A.P. Halestrap (1993) Am. J. Physiol. 264:C761-C782; Price, N.T. et al. (1998) Biochem. J. 329:321-328; and Martinelle, K. and I. Haggstrom (1993) J. Biotechnol. 30:339-350).

ATP-binding cassette (ABC) transporters are members of a superfamily of membrane proteins that transport substances ranging from small molecules such as ions, sugars, amino acids, peptides, and phospholipids, to lipopeptides, large proteins, and complex hydrophobic drugs. ABC transporters consist of four modules: two nucleotide-binding domains (NBD), which hydrolyze ATP to supply the energy required for transport, and two membrane-spanning domains (MSD), each containing six putative transmembrane segments. These four modules may be encoded by a single gene, as is the case for the cystic fibrosis transmembrane regulator (CFTR), or by separate genes. When encoded by separate genes, each gene product contains a single NBD and MSD. These "half-molecules" form homo- and heterodimers, such as Tap1 and Tap2, the endoplasmic reticulum-based major histocompatibility (MHC) peptide transport system. Several genetic diseases are attributed to defects in ABC transporters, such as the following diseases and their corresponding proteins: cystic fibrosis (CFTR, an ion channel), adrenoleukodystrophy (adrenoleukodystrophy protein, ALDP), Zellweger syndrome (peroxisomal membrane protein-70, PMP70), and hyperinsulinemic hypoglycemia (sulfonylurea receptor, SUR). Overexpression of the multidrug resistance (MDR) protein, another ABC transporter, in human cancer cells makes the cells resistant to a variety of cytotoxic drugs used in chemotherapy (Taglicht, D. and S. Michaelis (1998) Meth. Enzymol. 292:130-162).

A number of metal ions such as iron, zinc, copper, cobalt, manganese, molybdenum, selenium, nickel, and chromium are important as cofactors for a number of enzymes. For example, copper is involved in hemoglobin synthesis, connective tissue metabolism, and bone development, by acting as a cofactor in oxidoreductases such as superoxide dismutase, ferroxidase (ceruloplasmin), and lysyl oxidase. Copper and other metal ions must be provided in the diet, and are absorbed by transporters in the gastrointestinal tract. Plasma proteins transport the metal ions to the liver and other target organs, where specific transporters move the ions into cells and cellular organelles as needed. Imbalances in metal ion metabolism have been associated with a number of disease states (Danks, D.M. (1986) J. Med. Genet. 23:99-106).

Transport of fatty acids across the plasma membrane can occur by diffusion, a high capacity, low affinity process. However, under normal physiological conditions a significant fraction of fatty acid transport appears to occur via a high affinity, low capacity protein-mediated transport process.

Fatty acid transport protein (FATP), an integral membrane protein with four transmembrane segments, is expressed in tissues exhibiting high levels of plasma membrane fatty acid flux, such as muscle, heart, and adipose. Expression of FATP is upregulated in 3T3-L1 cells during adipose conversion, and expression in COS7 fibroblasts elevates uptake of long-chain fatty acids (Hui, T.Y. et al. (1998) J. Biol. Chem. 273:27420-27429).

Mitochondrial carrier proteins are transmembrane-spanning proteins which transport ions and charged metabolites between the cytosol and the mitochondrial matrix. Examples include the ADP, ATP carrier protein; the 2-oxoglutarate/malate carrier; the phosphate carrier protein; the pyruvate carrier; the dicarboxylate carrier which transports malate, succinate, fumarate, and phosphate; the tricarboxylate carrier which transports citrate and malate; and the Grave's disease carrier protein, a protein recognized by IgG in patients with active Grave's disease, an autoimmune disorder resulting in hyperthyroidism. Proteins in this family consist of three tandem repeats of an approximately 100 amino acid domain, each of which contains two transmembrane regions (Stryer, L. (1995) Biochemistry, W.H. Freeman and Company, New York NY, p. 551; PROSITE PDOC00189 Mitochondrial energy transfer proteins signature; Online Mendelian Inheritance in Man (OMIM) *275000 Graves Disease).

This class of transporters also includes the mitochondrial uncoupling proteins, which create proton leaks across the inner mitochondrial membrane, thus uncoupling oxidative phosphorylation from ATP synthesis. The result is energy dissipation in the form of heat. Mitochondrial uncoupling proteins have been implicated as modulators of thermoregulation and metabolic rate, and have been proposed as potential targets for drugs against metabolic diseases such as obesity (Ricquier, D. et al. (1999) J. Int. Med. 245:637-642).

Ion Channels

The electrical potential of a cell is generated and maintained by controlling the movement of ions across the plasma membrane. The movement of ions requires ion channels, which form ion-selective pores within the membrane. There are two basic types of ion channels, ion transporters and gated ion channels. Ion transporters utilize the energy obtained from ATP hydrolysis to actively transport an ion against the ion's concentration gradient. Gated ion channels allow passive flow of an ion down the ion's electrochemical gradient under restricted conditions. Together, these types of ion channels generate, maintain, and utilize an electrochemical gradient that is used in 1) electrical impulse conduction down the axon of a nerve cell, 2) transport of molecules into cells against concentration gradients, 3) initiation of muscle contraction, and 4) endocrine cell secretion.

Ion Transporters

Ion transporters generate and maintain the resting electrical potential of a cell. Utilizing the energy derived from ATP hydrolysis, they transport ions against the ion's concentration gradient.

These transmembrane ATPases are divided into three families. The phosphorylated (P) class ion transporters, including Na^+-K^+ ATPase, Ca^{2+}-ATPase, and H^+-ATPase, are activated by a phosphorylation event. P-class ion transporters are responsible for maintaining resting potential distributions such that cytosolic concentrations of Na^+ and Ca^{2+} are low and cytosolic concentration of K^+ is high. The vacuolar (V) class of ion transporters includes H^+ pumps on intracellular organelles, such as lysosomes and Golgi. V-class ion transporters are responsible for generating the low pH within the lumen of these organelles that is required for function. The coupling factor (F) class consists of H^+ pumps in the mitochondria. F-class ion transporters utilize a proton gradient to generate ATP from ADP and inorganic phosphate (P_i).

The P-ATPases are hexamers of a 100 kD subunit with ten transmembrane domains and several large cytoplasmic regions that may play a role in ion binding (Scarborough, G.A. (1999) Curr. Opin. Cell Biol. 11:517-522). The V-ATPases are composed of two functional domains: the V_1 domain, a peripheral complex responsible for ATP hydrolysis; and the V_0 domain, an integral complex responsible for proton translocation across the membrane. The F-ATPases are structurally and evolutionarily related to the V-ATPases. The F-ATPase F_0 domain contains 12 copies of the c subunit, a highly hydrophobic protein composed of two transmembrane domains and containing a single buried carboxyl group in TM2 that is essential for proton transport. The V-ATPase V_0 domain contains three types of homologous c subunits with four or five transmembrane domains and the essential carboxyl group in TM4 or TM3. Both types of complex also contain a single a subunit that may be involved in regulating the pH dependence of activity (Forgac, M. (1999) J. Biol. Chem. 274:12951-12954).

The resting potential of the cell is utilized in many processes involving carrier proteins and gated ion channels. Carrier proteins utilize the resting potential to transport molecules into and out of the cell. Amino acid and glucose transport into many cells is linked to sodium ion co-transport (symport) so that the movement of Na^+ down an electrochemical gradient drives transport of the other molecule up a concentration gradient. Similarly, cardiac muscle links transfer of Ca^{2+} out of the cell with transport of Na^+ into the cell (antiport).

Gated Ion Channels

Gated ion channels control ion flow by regulating the opening and closing of pores. The ability to control ion flux through various gating mechanisms allows ion channels to mediate such diverse signaling and homeostatic functions as neuronal and endocrine signaling, muscle contraction, fertilization, and regulation of ion and pH balance. Gated ion channels are categorized according to the manner of regulating the gating function. Mechanically-gated channels open their pores in response to mechanical stress; voltage-gated channels (e.g., Na^+, K^+, Ca^{2+}, and Cl^- channels) open their pores in response to changes in membrane potential; and ligand-gated channels (e.g.,

acetylcholine-, serotonin-, and glutamate-gated cation channels, and GABA- and glycine-gated chloride channels) open their pores in the presence of a specific ion, nucleotide, or neurotransmitter. The gating properties of a particular ion channel (i.e., its threshold for and duration of opening and closing) are sometimes modulated by association with auxiliary channel proteins and/or post translational modifications, such as phosphorylation.

Mechanically-gated or mechanosensitive ion channels act as transducers for the senses of touch, hearing, and balance, and also play important roles in cell volume regulation, smooth muscle contraction, and cardiac rhythm generation. A stretch-inactivated channel (SIC) was recently cloned from rat kidney. The SIC channel belongs to a group of channels which are activated by pressure or stress on the cell membrane and conduct both Ca^{2+} and Na^{+} (Suzuki, M. et al. (1999) J. Biol. Chem. 274:6330-6335).

The pore-forming subunits of the voltage-gated cation channels form a superfamily of ion channel proteins. The characteristic domain of these channel proteins comprises six transmembrane domains (S1-S6), a pore-forming region (P) located between S5 and S6, and intracellular amino and carboxy termini. In the Na^{+} and Ca^{2+} subfamilies, this domain is repeated four times, while in the K^{+} channel subfamily, each channel is formed from a tetramer of either identical or dissimilar subunits. The P region contains information specifying the ion selectivity for the channel. In the case of K^{+} channels, a GYG tripeptide is involved in this selectivity (Ishii, T.M. et al. (1997) Proc. Natl. Acad. Sci. USA 94:11651-11656).

Voltage-gated Na^{+} and K^{+} channels are necessary for the function of electrically excitable cells, such as nerve and muscle cells. Action potentials, which lead to neurotransmitter release and muscle contraction, arise from large, transient changes in the permeability of the membrane to Na^{+} and K^{+} ions. Depolarization of the membrane beyond the threshold level opens voltage-gated Na^{+} channels. Sodium ions flow into the cell, further depolarizing the membrane and opening more voltage-gated Na^{+} channels, which propagates the depolarization down the length of the cell. Depolarization also opens voltage-gated potassium channels. Consequently, potassium ions flow outward, which leads to repolarization of the membrane. Voltage-gated channels utilize charged residues in the fourth transmembrane segment (S4) to sense voltage change. The open state lasts only about 1 millisecond, at which time the channel spontaneously converts into an inactive state that cannot be opened irrespective of the membrane potential. Inactivation is mediated by the channel's N-terminus, which acts as a plug that closes the pore. The transition from an inactive to a closed state requires a return to resting potential.

Voltage-gated Na^{+} channels are heterotrimeric complexes composed of a 260 kDa pore-forming α subunit that associates with two smaller auxiliary subunits, β1 and β2. The β2 subunit is a integral membrane glycoprotein that contains an extracellular Ig domain, and its association with α

and β1 subunits correlates with increased functional expression of the channel, a change in its gating properties, as well as an increase in whole cell capacitance due to an increase in membrane surface area (Isom, L.L. et al. (1995) Cell 83:433-442).

Non voltage-gated Na^+ channels include the members of the amiloride-sensitive Na^+ channel/degenerin (NaC/DEG) family. Channel subunits of this family are thought to consist of two transmembrane domains flanking a long extracellular loop, with the amino and carboxyl termini located within the cell. The NaC/DEG family includes the epithelial Na^+ channel (ENaC) involved in Na^+ reabsorption in epithelia including the airway, distal colon, cortical collecting duct of the kidney, and exocrine duct glands. Mutations in ENaC result in pseudohypoaldosteronism type 1 and Liddle's syndrome (pseudohyperaldosteronism). The NaC/DEG family also includes the recently characterized H^+-gated cation channels or acid-sensing ion channels (ASIC). ASIC subunits are expressed in the brain and form heteromultimeric Na^+-permeable channels. These channels require acid pH fluctuations for activation. ASIC subunits show homology to the degenerins, a family of mechanically-gated channels originally isolated from C. elegans. Mutations in the degenerins cause neurodegeneration. ASIC subunits may also have a role in neuronal function, or in pain perception, since tissue acidosis causes pain (Waldmann, R. and M. Lazdunski (1998) Curr. Opin. Neurobiol. 8:418-424; Eglen, R.M. et al. (1999) Trends Pharmacol. Sci. 20:337-342).

K^+ channels are located in all cell types, and may be regulated by voltage, ATP concentration, or second messengers such as Ca^{2+} and cAMP. In non-excitable tissue, K^+ channels are involved in protein synthesis, control of endocrine secretions, and the maintenance of osmotic equilibrium across membranes. In neurons and other excitable cells, in addition to regulating action potentials and repolarizing membranes, K^+ channels are responsible for setting resting membrane potential. The cytosol contains non-diffusible anions and, to balance this net negative charge, the cell contains a Na^+-K^+ pump and ion channels that provide the redistribution of Na^+, K^+, and Cl^-. The pump actively transports Na^+ out of the cell and K^+ into the cell in a 3:2 ratio. Ion channels in the plasma membrane allow K^+ and Cl^- to flow by passive diffusion. Because of the high negative charge within the cytosol, Cl^- flows out of the cell. The flow of K^+ is balanced by an electromotive force pulling K^+ into the cell, and a K^+ concentration gradient pushing K^+ out of the cell. Thus, the resting membrane potential is primarily regulated by K^+ flow (Salkoff, L. and T. Jegla (1995) Neuron 15:489-492).

Potassium channel subunits of the Shaker-like superfamily all have the characteristic six transmembrane/1 pore domain structure. Four subunits combine as homo- or heterotetramers to form functional K channels. These pore-forming subunits also associate with various cytoplasmic β subunits that alter channel inactivation kinetics. The Shaker-like channel family includes the voltage-gated K^+ channels as well as the delayed rectifier type channels such as the human ether-a-go-go

related gene (HERG) associated with long QT, a cardiac dysrythmia syndrome (Curran, M.E. (1998) Curr. Opin. Biotechnol. 9:565-572; Kaczorowski, G.J. and M.L. Garcia (1999) Curr. Opin. Chem. Biol. 3:448-458).

A second superfamily of K^+ channels is composed of the inward rectifying channels (Kir). Kir channels have the property of preferentially conducting K^+ currents in the inward direction. These proteins consist of a single potassium selective pore domain and two transmembrane domains, which correspond to the fifth and sixth transmembrane domains of voltage-gated K^+ channels. Kir subunits also associate as tetramers. The Kir family includes ROMK1, mutations in which lead to Bartter syndrome, a renal tubular disorder. Kir channels are also involved in regulation of cardiac pacemaker activity, seizures and epilepsy, and insulin regulation (Doupnik, C.A. et al. (1995) Curr. Opin. Neurobiol. 5:268-277; Curran, supra).

The recently recognized TWIK K^+ channel family includes the mammalian TWIK-1, TREK-1 and TASK proteins. Members of this family possess an overall structure with four transmembrane domains and two P domains. These proteins are probably involved in controlling the resting potential in a large set of cell types (Duprat, F. et al. (1997) EMBO J 16:5464-5471).

The voltage-gated Ca^{2+} channels have been classified into several subtypes based upon their electrophysiological and pharmacological characteristics. L-type Ca^{2+} channels are predominantly expressed in heart and skeletal muscle where they play an essential role in excitation-contraction coupling. T-type channels are important for cardiac pacemaker activity, while N-type and P/Q-type channels are involved in the control of neurotransmitter release in the central and peripheral nervous system. The L-type and N-type voltage-gated Ca^{2+} channels have been purified and, though their functions differ dramatically, they have similar subunit compositions. The channels are composed of three subunits. The α_1 subunit forms the membrane pore and voltage sensor, while the $\alpha_2\delta$ and β subunits modulate the voltage-dependence, gating properties, and the current amplitude of the channel. These subunits are encoded by at least six α_1, one $\alpha_2\delta$, and four β genes. A fourth subunit, γ, has been identified in skeletal muscle (Walker, D. et al. (1998) J. Biol. Chem. 273:2361-2367; McCleskey, E.W. (1994) Curr. Opin. Neurobiol. 4:304-312).

The transient receptor family (Trp) of calcium ion channels are thought to mediate capacitative calcium entry (CCE). CCE is the Ca^{2+} influx into cells to resupply Ca^{2+} stores depleted by the action of inositol triphosphate (IP3) and other agents in response to numerous hormones and growth factors. Trp and Trp-like were first cloned from Drosophila and have similarity to voltage gated Ca2+ channels in the S3 through S6 regions. This suggests that Trp and/or related proteins may form mammalian CCC entry channels (Zhu, X. et al. (1996) Cell 85:661-671; Boulay, G. et al. (1997) J. Biol. Chem. 272:29672-29680). Melastatin is a gene isolated in both the mouse and human, and whose expression in melanoma cells is inversely correlated with melanoma aggressiveness in vivo.

The human cDNA transcript corresponds to a 1533-amino acid protein having homology to members of the Trp family. It has been proposed that the combined use of malastatin mRNA expression status and tumor thickness might allow for the determination of subgroups of patients at both low and high risk for developing metastatic disease (Duncan, L.M. et al (2001) J. Clin. Oncol. 19:568-576).

Chloride channels are necessary in endocrine secretion and in regulation of cytosolic and organelle pH. In secretory epithelial cells, Cl^- enters the cell across a basolateral membrane through an Na^+, K^+/Cl^- cotransporter, accumulating in the cell above its electrochemical equilibrium concentration. Secretion of Cl^- from the apical surface, in response to hormonal stimulation, leads to flow of Na^+ and water into the secretory lumen. The cystic fibrosis transmembrane conductance regulator (CFTR) is a chloride channel encoded by the gene for cystic fibrosis, a common fatal genetic disorder in humans. CFTR is a member of the ABC transporter family, and is composed of two domains each consisting of six transmembrane domains followed by a nucleotide-binding site. Loss of CFTR function decreases transepithelial water secretion and, as a result, the layers of mucus that coat the respiratory tree, pancreatic ducts, and intestine are dehydrated and difficult to clear. The resulting blockage of these sites leads to pancreatic insufficiency, "meconium ileus", and devastating "chronic obstructive pulmonary disease" (Al-Awqati, Q. et al. (1992) J. Exp. Biol. 172:245-266).

The voltage-gated chloride channels (CLC) are characterized by 10-12 transmembrane domains, as well as two small globular domains known as CBS domains. The CLC subunits probably function as homotetramers. CLC proteins are involved in regulation of cell volume, membrane potential stabilization, signal transduction, and transepithelial transport. Mutations in CLC-1, expressed predominantly in skeletal muscle, are responsible for autosomal recessive generalized myotonia and autosomal dominant myotonia congenita, while mutations in the kidney channel CLC-5 lead to kidney stones (Jentsch, T.J. (1996) Curr. Opin. Neurobiol. 6:303-310).

Ligand-gated channels open their pores when an extracellular or intracellular mediator binds to the channel. Neurotransmitter-gated channels are channels that open when a neurotransmitter binds to their extracellular domain. These channels exist in the postsynaptic membrane of nerve or muscle cells. There are two types of neurotransmitter-gated channels. Sodium channels open in response to excitatory neurotransmitters, such as acetylcholine, glutamate, and serotonin. This opening causes an influx of Na^+ and produces the initial localized depolarization that activates the voltage-gated channels and starts the action potential. Chloride channels open in response to inhibitory neurotransmitters, such as γ-aminobutyric acid (GABA) and glycine, leading to hyperpolarization of the membrane and the subsequent generation of an action potential. Neurotransmitter-gated ion channels have four transmembrane domains and probably function as pentamers (Jentsch, supra). Amino acids in the second transmembrane domain appear to be important in determining channel permeation and selectivity (Sather, W.A. et al. (1994) Curr. Opin. Neurobiol.

4:313-323).

Ligand-gated channels can be regulated by intracellular second messengers. For example, calcium-activated K^+ channels are gated by internal calcium ions. In nerve cells, an influx of calcium during depolarization opens K^+ channels to modulate the magnitude of the action potential (Ishi et al., supra). The large conductance (BK) channel has been purified from brain and its subunit composition determined. The α subunit of the BK channel has seven rather than six transmembrane domains in contrast to voltage-gated K^+ channels. The extra transmembrane domain is located at the subunit N-terminus. A 28-amino-acid stretch in the C-terminal region of the subunit (the "calcium bowl" region) contains many negatively charged residues and is thought to be the region responsible for calcium binding. The β subunit consists of two transmembrane domains connected by a glycosylated extracellular loop, with intracellular N- and C-termini (Kaczorowski, supra; Vergara, C. et al. (1998) Curr. Opin. Neurobiol. 8:321-329).

Cyclic nucleotide-gated (CNG) channels are gated by cytosolic cyclic nucleotides. The best examples of these are the cAMP-gated Na^+ channels involved in olfaction and the cGMP-gated cation channels involved in vision. Both systems involve ligand-mediated activation of a G-protein coupled receptor which then alters the level of cyclic nucleotide within the cell. CNG channels also represent a major pathway for Ca^{2+} entry into neurons, and play roles in neuronal development and plasticity. CNG channels are tetramers containing at least two types of subunits, an α subunit which can form functional homomeric channels, and a β subunit, which modulates the channel properties. All CNG subunits have six transmembrane domains and a pore forming region between the fifth and sixth transmembrane domains, similar to voltage-gated K^+ channels. A large C-terminal domain contains a cyclic nucleotide binding domain, while the N-terminal domain confers variation among channel subtypes (Zufall, F. et al. (1997) Curr. Opin. Neurobiol. 7:404-412).

The activity of other types of ion channel proteins may also be modulated by a variety of intracellular signalling proteins. Many channels have sites for phosphorylation by one or more protein kinases including protein kinase A, protein kinase C, tyrosine kinase, and casein kinase II, all of which regulate ion channel activity in cells. Kir channels are activated by the binding of the $G\beta\gamma$ subunits of heterotrimeric G-proteins (Reimann, F. and F.M. Ashcroft (1999) Curr. Opin. Cell. Biol. 11:503-508). Other proteins are involved in the localization of ion channels to specific sites in the cell membrane. Such proteins include the PDZ domain proteins known as MAGUKs (membrane-associated guanylate kinases) which regulate the clustering of ion channels at neuronal synapses (Craven, S.E. and D.S. Bredt (1998) Cell 93:495-498).

Disease Correlation

The etiology of numerous human diseases and disorders can be attributed to defects in the

transport of molecules across membranes. Defects in the trafficking of membrane-bound transporters and ion channels are associated with several disorders, e.g., cystic fibrosis, glucose-galactose malabsorption syndrome, hypercholesterolemia, von Gierke disease, and certain forms of diabetes mellitus. Single-gene defect diseases resulting in an inability to transport small molecules across membranes include, e.g., cystinuria, iminoglycinuria, Hartup disease, and Fanconi disease (van't Hoff, W.G. (1996) Exp. Nephrol. 4:253-262; Talente, G.M. et al. (1994) Ann. Intern. Med. 120:218-226; and Chillon, M. et al. (1995) New Engl. J. Med. 332:1475-1480).

Human diseases caused by mutations in ion channel genes include disorders of skeletal muscle, cardiac muscle, and the central nervous system. Mutations in the pore-forming subunits of sodium and chloride channels cause myotonia, a muscle disorder in which relaxation after voluntary contraction is delayed. Sodium channel myotonias have been treated with channel blockers. Mutations in muscle sodium and calcium channels cause forms of periodic paralysis, while mutations in the sarcoplasmic calcium release channel, T-tubule calcium channel, and muscle sodium channel cause malignant hyperthermia. Cardiac arrythmia disorders such as the long QT syndromes and idiopathic ventricular fibrillation are caused by mutations in potassium and sodium channels (Cooper, E.C. and L.Y. Jan (1998) Proc. Natl. Acad. Sci. USA 96:4759-4766). All four known human idiopathic epilepsy genes code for ion channel proteins (Berkovic, S.F. and I.E. Scheffer (1999) Curr. Opin. Neurology 12:177-182). Other neurological disorders such as ataxias, hemiplegic migraine and hereditary deafness can also result from mutations in ion channel genes (Jen, J. (1999) Curr. Opin. Neurobiol. 9:274-280; Cooper, supra).

Ion channels have been the target for many drug therapies. Neurotransmitter-gated channels have been targeted in therapies for treatment of insomnia, anxiety, depression, and schizophrenia. Voltage-gated channels have been targeted in therapies for arrhythmia, ischemic stroke, head trauma, and neurodegenerative disease (Taylor, C.P. and L.S. Narasimhan (1997) Adv. Pharmacol. 39:47-98). Various classes of ion channels also play an important role in the perception of pain, and thus are potential targets for new analgesics. These include the vanilloid-gated ion channels, which are activated by the vanilloid capsaicin, as well as by noxious heat. Local anesthetics such as lidocaine and mexiletine which blockade voltage-gated Na^+ channels have been useful in the treatment of neuropathic pain (Eglen, supra).

Ion channels in the immune system have recently been suggested as targets for immunomodulation. T-cell activation depends upon calcium signaling, and a diverse set of T-cell specific ion channels has been characterized that affect this signaling process. Channel blocking agents can inhibit secretion of lymphokines, cell proliferation, and killing of target cells. A peptide antagonist of the T-cell potassium channel Kv1.3 was found to suppress delayed-type hypersensitivity and allogenic responses in pigs, validating the idea of channel blockers as safe and efficacious

immunosuppressants (Cahalan, M.D. and K.G. Chandy (1997) Curr. Opin. Biotechnol. 8:749-756).

The discovery of new transporters and ion channels, and the polynucleotides encoding them, satisfies a need in the art by providing new compositions which are useful in the diagnosis, prevention, and treatment of transport, neurological, muscle, immunological, and cell proliferative disorders, and in the assessment of the effects of exogenous compounds on the expression of nucleic acid and amino acid sequences of transporters and ion channels.

SUMMARY OF THE INVENTION

The invention features purified polypeptides, transporters and ion channels, referred to collectively as "TRICH" and individually as "TRICH-1," "TRICH-2," "TRICH-3," "TRICH-4," "TRICH-5," "TRICH-6," "TRICH-7," "TRICH-8," "TRICH-9," "TRICH-10," "TRICH-11," "TRICH-12," "TRICH-13," "TRICH-14," "TRICH-15," "TRICH-16," "TRICH-17," "TRICH-18," "TRICH-19," "TRICH-20," "TRICH-21," "TRICH-22," "TRICH-23," "TRICH-24," "TRICH-25," "TRICH-26," "TRICH-27," "TRICH-28," "TRICH-29," and "TRICH-30." In one aspect, the invention provides an isolated polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-30, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-30, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-30, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-30. In one alternative, the invention provides an isolated polypeptide comprising the amino acid sequence of SEQ ID NO:1-30.

The invention further provides an isolated polynucleotide encoding a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-30, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-30, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-30, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-30. In one alternative, the polynucleotide encodes a polypeptide selected from the group consisting of SEQ ID NO:1-30. In another alternative, the polynucleotide is selected from the group consisting of SEQ ID NO:31-60.

Additionally, the invention provides a recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide encoding a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting

of SEQ ID NO:1-30, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-30, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-30, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-30. In one alternative, the invention provides a cell transformed with the recombinant polynucleotide. In another alternative, the invention provides a transgenic organism comprising the recombinant polynucleotide.

The invention also provides a method for producing a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-30, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-30, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-30, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-30. The method comprises a) culturing a cell under conditions suitable for expression of the polypeptide, wherein said cell is transformed with a recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide encoding the polypeptide, and b) recovering the polypeptide so expressed.

Additionally, the invention provides an isolated antibody which specifically binds to a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-30, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-30, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-30, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-30.

The invention further provides an isolated polynucleotide selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:31-60, b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:31-60, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). In one alternative, the polynucleotide comprises at least 60 contiguous nucleotides.

Additionally, the invention provides a method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group

consisting of SEQ ID NO:31-60, b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:31-60, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). The method comprises a) hybridizing the sample with a probe comprising at least 20 contiguous nucleotides comprising a sequence complementary to said target polynucleotide in the sample, and which probe specifically hybridizes to said target polynucleotide, under conditions whereby a hybridization complex is formed between said probe and said target polynucleotide or fragments thereof, and b) detecting the presence or absence of said hybridization complex, and optionally, if present, the amount thereof. In one alternative, the probe comprises at least 60 contiguous nucleotides.

The invention further provides a method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:31-60, b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:31-60, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). The method comprises a) amplifying said target polynucleotide or fragment thereof using polymerase chain reaction amplification, and b) detecting the presence or absence of said amplified target polynucleotide or fragment thereof, and, optionally, if present, the amount thereof.

The invention further provides a composition comprising an effective amount of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-30, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-30, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-30, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-30, and a pharmaceutically acceptable excipient. In one embodiment, the composition comprises an amino acid sequence selected from the group consisting of SEQ ID NO:1-30. The invention additionally provides a method of treating a disease or condition associated with decreased expression of functional TRICH, comprising administering to a patient in need of such treatment the composition.

The invention also provides a method for screening a compound for effectiveness as an agonist of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino

acid sequence selected from the group consisting of SEQ ID NO:1-30, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-30, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-30, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-30. The method comprises a) exposing a sample comprising the polypeptide to a compound, and b) detecting agonist activity in the sample. In one alternative, the invention provides a composition comprising an agonist compound identified by the method and a pharmaceutically acceptable excipient. In another alternative, the invention provides a method of treating a disease or condition associated with decreased expression of functional TRICH, comprising administering to a patient in need of such treatment the composition.

Additionally, the invention provides a method for screening a compound for effectiveness as an antagonist of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-30, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-30, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-30, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-30. The method comprises a) exposing a sample comprising the polypeptide to a compound, and b) detecting antagonist activity in the sample. In one alternative, the invention provides a composition comprising an antagonist compound identified by the method and a pharmaceutically acceptable excipient. In another alternative, the invention provides a method of treating a disease or condition associated with overexpression of functional TRICH, comprising administering to a patient in need of such treatment the composition.

The invention further provides a method of screening for a compound that specifically binds to a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-30, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-30, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-30, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-30. The method comprises a) combining the polypeptide with at least one test compound under suitable conditions, and b) detecting binding of the polypeptide to the test compound, thereby identifying a compound that specifically binds to the polypeptide.

The invention further provides a method of screening for a compound that modulates the activity of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-30, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-30, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-30, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-30. The method comprises a) combining the polypeptide with at least one test compound under conditions permissive for the activity of the polypeptide, b) assessing the activity of the polypeptide in the presence of the test compound, and c) comparing the activity of the polypeptide in the presence of the test compound with the activity of the polypeptide in the absence of the test compound, wherein a change in the activity of the polypeptide in the presence of the test compound is indicative of a compound that modulates the activity of the polypeptide.

The invention further provides a method for screening a compound for effectiveness in altering expression of a target polynucleotide, wherein said target polynucleotide comprises a polynucleotide sequence selected from the group consisting of SEQ ID NO:31-60, the method comprising a) exposing a sample comprising the target polynucleotide to a compound, and b) detecting altered expression of the target polynucleotide.

The invention further provides a method for assessing toxicity of a test compound, said method comprising a) treating a biological sample containing nucleic acids with the test compound; b) hybridizing the nucleic acids of the treated biological sample with a probe comprising at least 20 contiguous nucleotides of a polynucleotide selected from the group consisting of i) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:31-60, ii) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:31-60, iii) a polynucleotide having a sequence complementary to i), iv) a polynucleotide complementary to the polynucleotide of ii), and v) an RNA equivalent of i)-iv). Hybridization occurs under conditions whereby a specific hybridization complex is formed between said probe and a target polynucleotide in the biological sample, said target polynucleotide selected from the group consisting of i) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:31-60, ii) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:31-60, iii) a polynucleotide complementary to the polynucleotide of i), iv) a polynucleotide complementary to the polynucleotide of ii), and v) an RNA equivalent of i)-iv). Alternatively, the target

polynucleotide comprises a fragment of a polynucleotide sequence selected from the group consisting of i)-v) above; c) quantifying the amount of hybridization complex; and d) comparing the amount of hybridization complex in the treated biological sample with the amount of hybridization complex in an untreated biological sample, wherein a difference in the amount of hybridization complex in the treated biological sample is indicative of toxicity of the test compound.

BRIEF DESCRIPTION OF THE TABLES

Table 1 summarizes the nomenclature for the full length polynucleotide and polypeptide sequences of the present invention.

Table 2 shows the GenBank identification number and annotation of the nearest GenBank homolog for polypeptides of the invention. The probability score for the match between each polypeptide and its GenBank homolog is also shown.

Table 3 shows structural features of polypeptide sequences of the invention, including predicted motifs and domains, along with the methods, algorithms, and searchable databases used for analysis of the polypeptides.

Table 4 lists the cDNA and/or genomic DNA fragments which were used to assemble polynucleotide sequences of the invention, along with selected fragments of the polynucleotide sequences.

Table 5 shows the representative cDNA library for polynucleotides of the invention.

Table 6 provides an appendix which describes the tissues and vectors used for construction of the cDNA libraries shown in Table 5.

Table 7 shows the tools, programs, and algorithms used to analyze the polynucleotides and polypeptides of the invention, along with applicable descriptions, references, and threshold parameters.

DESCRIPTION OF THE INVENTION

Before the present proteins, nucleotide sequences, and methods are described, it is understood that this invention is not limited to the particular machines, materials and methods described, as these may vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only, and is not intended to limit the scope of the present invention which will be limited only by the appended claims.

It must be noted that as used herein and in the appended claims, the singular forms "a," "an," and "the" include plural reference unless the context clearly dictates otherwise. Thus, for example, a reference to "a host cell" includes a plurality of such host cells, and a reference to "an antibody" is a reference to one or more antibodies and equivalents thereof known to those skilled in the art, and so

forth.

Unless defined otherwise, all technical and scientific terms used herein have the same meanings as commonly understood by one of ordinary skill in the art to which this invention belongs. Although any machines, materials, and methods similar or equivalent to those described herein can be used to practice or test the present invention, the preferred machines, materials and methods are now described. All publications mentioned herein are cited for the purpose of describing and disclosing the cell lines, protocols, reagents and vectors which are reported in the publications and which might be used in connection with the invention. Nothing herein is to be construed as an admission that the invention is not entitled to antedate such disclosure by virtue of prior invention.

DEFINITIONS

"TRICH" refers to the amino acid sequences of substantially purified TRICH obtained from any species, particularly a mammalian species, including bovine, ovine, porcine, murine, equine, and human, and from any source, whether natural, synthetic, semi-synthetic, or recombinant.

The term "agonist" refers to a molecule which intensifies or mimics the biological activity of TRICH. Agonists may include proteins, nucleic acids, carbohydrates, small molecules, or any other compound or composition which modulates the activity of TRICH either by directly interacting with TRICH or by acting on components of the biological pathway in which TRICH participates.

An "allelic variant" is an alternative form of the gene encoding TRICH. Allelic variants may result from at least one mutation in the nucleic acid sequence and may result in altered mRNAs or in polypeptides whose structure or function may or may not be altered. A gene may have none, one, or many allelic variants of its naturally occurring form. Common mutational changes which give rise to allelic variants are generally ascribed to natural deletions, additions, or substitutions of nucleotides. Each of these types of changes may occur alone, or in combination with the others, one or more times in a given sequence.

"Altered" nucleic acid sequences encoding TRICH include those sequences with deletions, insertions, or substitutions of different nucleotides, resulting in a polypeptide the same as TRICH or a polypeptide with at least one functional characteristic of TRICH. Included within this definition are polymorphisms which may or may not be readily detectable using a particular oligonucleotide probe of the polynucleotide encoding TRICH, and improper or unexpected hybridization to allelic variants, with a locus other than the normal chromosomal locus for the polynucleotide sequence encoding TRICH. The encoded protein may also be "altered," and may contain deletions, insertions, or substitutions of amino acid residues which produce a silent change and result in a functionally equivalent TRICH. Deliberate amino acid substitutions may be made on the basis of similarity in polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues, as long as the biological or immunological activity of TRICH is retained. For example,

negatively charged amino acids may include aspartic acid and glutamic acid, and positively charged amino acids may include lysine and arginine. Amino acids with uncharged polar side chains having similar hydrophilicity values may include: asparagine and glutamine; and serine and threonine. Amino acids with uncharged side chains having similar hydrophilicity values may include: leucine, isoleucine, and valine; glycine and alanine; and phenylalanine and tyrosine.

The terms "amino acid" and "amino acid sequence" refer to an oligopeptide, peptide, polypeptide, or protein sequence, or a fragment of any of these, and to naturally occurring or synthetic molecules. Where "amino acid sequence" is recited to refer to a sequence of a naturally occurring protein molecule, "amino acid sequence" and like terms are not meant to limit the amino acid sequence to the complete native amino acid sequence associated with the recited protein molecule.

"Amplification" relates to the production of additional copies of a nucleic acid sequence. Amplification is generally carried out using polymerase chain reaction (PCR) technologies well known in the art.

The term "antagonist" refers to a molecule which inhibits or attenuates the biological activity of TRICH. Antagonists may include proteins such as antibodies, nucleic acids, carbohydrates, small molecules, or any other compound or composition which modulates the activity of TRICH either by directly interacting with TRICH or by acting on components of the biological pathway in which TRICH participates.

The term "antibody" refers to intact immunoglobulin molecules as well as to fragments thereof, such as Fab, $F(ab')_2$, and Fv fragments, which are capable of binding an epitopic determinant. Antibodies that bind TRICH polypeptides can be prepared using intact polypeptides or using fragments containing small peptides of interest as the immunizing antigen. The polypeptide or oligopeptide used to immunize an animal (e.g., a mouse, a rat, or a rabbit) can be derived from the translation of RNA, or synthesized chemically, and can be conjugated to a carrier protein if desired. Commonly used carriers that are chemically coupled to peptides include bovine serum albumin, thyroglobulin, and keyhole limpet hemocyanin (KLH). The coupled peptide is then used to immunize the animal.

The term "antigenic determinant" refers to that region of a molecule (i.e., an epitope) that makes contact with a particular antibody. When a protein or a fragment of a protein is used to immunize a host animal, numerous regions of the protein may induce the production of antibodies which bind specifically to antigenic determinants (particular regions or three-dimensional structures on the protein). An antigenic determinant may compete with the intact antigen (i.e., the immunogen used to elicit the immune response) for binding to an antibody.

The term "antisense" refers to any composition capable of base-pairing with the "sense" (coding) strand of a specific nucleic acid sequence. Antisense compositions may include DNA;

RNA; peptide nucleic acid (PNA); oligonucleotides having modified backbone linkages such as phosphorothioates, methylphosphonates, or benzylphosphonates; oligonucleotides having modified sugar groups such as 2'-methoxyethyl sugars or 2'-methoxyethoxy sugars; or oligonucleotides having modified bases such as 5-methyl cytosine, 2'-deoxyuracil, or 7-deaza-2'-deoxyguanosine. Antisense molecules may be produced by any method including chemical synthesis or transcription. Once introduced into a cell, the complementary antisense molecule base-pairs with a naturally occurring nucleic acid sequence produced by the cell to form duplexes which block either transcription or translation. The designation "negative" or "minus" can refer to the antisense strand, and the designation "positive" or "plus" can refer to the sense strand of a reference DNA molecule.

The term "biologically active" refers to a protein having structural, regulatory, or biochemical functions of a naturally occurring molecule. Likewise, "immunologically active" or "immunogenic" refers to the capability of the natural, recombinant, or synthetic TRICH, or of any oligopeptide thereof, to induce a specific immune response in appropriate animals or cells and to bind with specific antibodies.

"Complementary" describes the relationship between two single-stranded nucleic acid sequences that anneal by base-pairing. For example, 5'-AGT-3' pairs with its complement, 3'-TCA-5'.

A "composition comprising a given polynucleotide sequence" and a "composition comprising a given amino acid sequence" refer broadly to any composition containing the given polynucleotide or amino acid sequence. The composition may comprise a dry formulation or an aqueous solution. Compositions comprising polynucleotide sequences encoding TRICH or fragments of TRICH may be employed as hybridization probes. The probes may be stored in freeze-dried form and may be associated with a stabilizing agent such as a carbohydrate. In hybridizations, the probe may be deployed in an aqueous solution containing salts (e.g., NaCl), detergents (e.g., sodium dodecyl sulfate; SDS), and other components (e.g., Denhardt's solution, dry milk, salmon sperm DNA, etc.).

"Consensus sequence" refers to a nucleic acid sequence which has been subjected to repeated DNA sequence analysis to resolve uncalled bases, extended using the XL-PCR kit (Applied Biosystems, Foster City CA) in the 5' and/or the 3' direction, and resequenced, or which has been assembled from one or more overlapping cDNA, EST, or genomic DNA fragments using a computer program for fragment assembly, such as the GELVIEW fragment assembly system (GCG, Madison WI) or Phrap (University of Washington, Seattle WA). Some sequences have been both extended and assembled to produce the consensus sequence.

"Conservative amino acid substitutions" are those substitutions that are predicted to least interfere with the properties of the original protein, i.e., the structure and especially the function of the protein is conserved and not significantly changed by such substitutions. The table below shows

amino acids which may be substituted for an original amino acid in a protein and which are regarded as conservative amino acid substitutions.

Original Residue	Conservative Substitution
Ala	Gly, Ser
Arg	His, Lys
Asn	Asp, Gln, His
Asp	Asn, Glu
Cys	Ala, Ser
Gln	Asn, Glu, His
Glu	Asp, Gln, His
Gly	Ala
His	Asn, Arg, Gln, Glu
Ile	Leu, Val
Leu	Ile, Val
Lys	Arg, Gln, Glu
Met	Leu, Ile
Phe	His, Met, Leu, Trp, Tyr
Ser	Cys, Thr
Thr	Ser, Val
Trp	Phe, Tyr
Tyr	His, Phe, Trp
Val	Ile, Leu, Thr

Conservative amino acid substitutions generally maintain (a) the structure of the polypeptide backbone in the area of the substitution, for example, as a beta sheet or alpha helical conformation, (b) the charge or hydrophobicity of the molecule at the site of the substitution, and/or (c) the bulk of the side chain.

A "deletion" refers to a change in the amino acid or nucleotide sequence that results in the absence of one or more amino acid residues or nucleotides.

The term "derivative" refers to a chemically modified polynucleotide or polypeptide. Chemical modifications of a polynucleotide can include, for example, replacement of hydrogen by an alkyl, acyl, hydroxyl, or amino group. A derivative polynucleotide encodes a polypeptide which retains at least one biological or immunological function of the natural molecule. A derivative polypeptide is one modified by glycosylation, pegylation, or any similar process that retains at least one biological or immunological function of the polypeptide from which it was derived.

A "detectable label" refers to a reporter molecule or enzyme that is capable of generating a measurable signal and is covalently or noncovalently joined to a polynucleotide or polypeptide.

"Differential expression" refers to increased or upregulated; or decreased, downregulated, or absent gene or protein expression, determined by comparing at least two different samples. Such comparisons may be carried out between, for example, a treated and an untreated sample, or a diseased and a normal sample.

"Exon shuffling" refers to the recombination of different coding regions (exons). Since an exon may represent a structural or functional domain of the encoded protein, new proteins may be assembled through the novel reassortment of stable substructures, thus allowing acceleration of the evolution of new protein functions.

A "fragment" is a unique portion of TRICH or the polynucleotide encoding TRICH which is identical in sequence to but shorter in length than the parent sequence. A fragment may comprise up to the entire length of the defined sequence, minus one nucleotide/amino acid residue. For example, a fragment may comprise from 5 to 1000 contiguous nucleotides or amino acid residues. A fragment used as a probe, primer, antigen, therapeutic molecule, or for other purposes, may be at least 5, 10, 15, 16, 20, 25, 30, 40, 50, 60, 75, 100, 150, 250 or at least 500 contiguous nucleotides or amino acid residues in length. Fragments may be preferentially selected from certain regions of a molecule. For example, a polypeptide fragment may comprise a certain length of contiguous amino acids selected from the first 250 or 500 amino acids (or first 25% or 50%) of a polypeptide as shown in a certain defined sequence. Clearly these lengths are exemplary, and any length that is supported by the specification, including the Sequence Listing, tables, and figures, may be encompassed by the present embodiments.

A fragment of SEQ ID NO:31-60 comprises a region of unique polynucleotide sequence that specifically identifies SEQ ID NO:31-60, for example, as distinct from any other sequence in the genome from which the fragment was obtained. A fragment of SEQ ID NO:31-60 is useful, for example, in hybridization and amplification technologies and in analogous methods that distinguish SEQ ID NO:31-60 from related polynucleotide sequences. The precise length of a fragment of SEQ ID NO:31-60 and the region of SEQ ID NO:31-60 to which the fragment corresponds are routinely determinable by one of ordinary skill in the art based on the intended purpose for the fragment.

A fragment of SEQ ID NO:1-30 is encoded by a fragment of SEQ ID NO:31-60. A fragment of SEQ ID NO:1-30 comprises a region of unique amino acid sequence that specifically identifies SEQ ID NO:1-30. For example, a fragment of SEQ ID NO:1-30 is useful as an immunogenic peptide for the development of antibodies that specifically recognize SEQ ID NO:1-30. The precise length of a fragment of SEQ ID NO:1-30 and the region of SEQ ID NO:1-30 to which the fragment corresponds are routinely determinable by one of ordinary skill in the art based on the intended purpose for the fragment.

A "full length" polynucleotide sequence is one containing at least a translation initiation codon (e.g., methionine) followed by an open reading frame and a translation termination codon. A "full length" polynucleotide sequence encodes a "full length" polypeptide sequence.

"Homology" refers to sequence similarity or, interchangeably, sequence identity, between two or more polynucleotide sequences or two or more polypeptide sequences.

The terms "percent identity" and "% identity," as applied to polynucleotide sequences, refer to the percentage of residue matches between at least two polynucleotide sequences aligned using a standardized algorithm. Such an algorithm may insert, in a standardized and reproducible way, gaps in the sequences being compared in order to optimize alignment between two sequences, and therefore achieve a more meaningful comparison of the two sequences.

Percent identity between polynucleotide sequences may be determined using the default parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e sequence alignment program. This program is part of the LASERGENE software package, a suite of molecular biological analysis programs (DNASTAR, Madison WI). CLUSTAL V is described in Higgins, D.G. and P.M. Sharp (1989) CABIOS 5:151-153 and in Higgins, D.G. et al. (1992) CABIOS 8:189-191. For pairwise alignments of polynucleotide sequences, the default parameters are set as follows: Ktuple=2, gap penalty=5, window=4, and "diagonals saved"=4. The "weighted" residue weight table is selected as the default. Percent identity is reported by CLUSTAL V as the "percent similarity" between aligned polynucleotide sequences.

Alternatively, a suite of commonly used and freely available sequence comparison algorithms is provided by the National Center for Biotechnology Information (NCBI) Basic Local Alignment Search Tool (BLAST) (Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410), which is available from several sources, including the NCBI, Bethesda, MD, and on the Internet at http://www.ncbi.nlm.nih.gov/BLAST/. The BLAST software suite includes various sequence analysis programs including "blastn," that is used to align a known polynucleotide sequence with other polynucleotide sequences from a variety of databases. Also available is a tool called "BLAST 2 Sequences" that is used for direct pairwise comparison of two nucleotide sequences. "BLAST 2 Sequences" can be accessed and used interactively at http://www.ncbi.nlm.nih.gov/gorf/bl2.html. The "BLAST 2 Sequences" tool can be used for both blastn and blastp (discussed below). BLAST programs are commonly used with gap and other parameters set to default settings. For example, to compare two nucleotide sequences, one may use blastn with the "BLAST 2 Sequences" tool Version 2.0.12 (April-21-2000) set at default parameters. Such default parameters may be, for example:

Matrix: BLOSUM62

Reward for match: 1

Penalty for mismatch: -2

Open Gap: 5 and Extension Gap: 2 penalties

Gap x drop-off: 50

Expect: 10

Word Size: 11

Filter: on

Percent identity may be measured over the length of an entire defined sequence, for example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for example, over the length of a fragment taken from a larger, defined sequence, for instance, a fragment of at least 20, at least 30, at least 40, at least 50, at least 70, at least 100, or at least 200 contiguous nucleotides. Such lengths are exemplary only, and it is understood that any fragment length supported by the sequences shown herein, in the tables, figures, or Sequence Listing, may be used to describe a length over which percentage identity may be measured.

Nucleic acid sequences that do not show a high degree of identity may nevertheless encode similar amino acid sequences due to the degeneracy of the genetic code. It is understood that changes in a nucleic acid sequence can be made using this degeneracy to produce multiple nucleic acid sequences that all encode substantially the same protein.

The phrases "percent identity" and "% identity," as applied to polypeptide sequences, refer to the percentage of residue matches between at least two polypeptide sequences aligned using a standardized algorithm. Methods of polypeptide sequence alignment are well-known. Some alignment methods take into account conservative amino acid substitutions. Such conservative substitutions, explained in more detail above, generally preserve the charge and hydrophobicity at the site of substitution, thus preserving the structure (and therefore function) of the polypeptide.

Percent identity between polypeptide sequences may be determined using the default parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e sequence alignment program (described and referenced above). For pairwise alignments of polypeptide sequences using CLUSTAL V, the default parameters are set as follows: Ktuple=1, gap penalty=3, window=5, and "diagonals saved"=5. The PAM250 matrix is selected as the default residue weight table. As with polynucleotide alignments, the percent identity is reported by CLUSTAL V as the "percent similarity" between aligned polypeptide sequence pairs.

Alternatively the NCBI BLAST software suite may be used. For example, for a pairwise comparison of two polypeptide sequences, one may use the "BLAST 2 Sequences" tool Version 2.0.12 (April-21-2000) with blastp set at default parameters. Such default parameters may be, for example:

Matrix: BLOSUM62

Open Gap: 11 and Extension Gap: 1 penalties

Gap x drop-off: 50

Expect: 10

Word Size: 3

Filter: on

Percent identity may be measured over the length of an entire defined polypeptide sequence,

for example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for example, over the length of a fragment taken from a larger, defined polypeptide sequence, for instance, a fragment of at least 15, at least 20, at least 30, at least 40, at least 50, at least 70 or at least 150 contiguous residues. Such lengths are exemplary only, and it is understood that any fragment length supported by the sequences shown herein, in the tables, figures or Sequence Listing, may be used to describe a length over which percentage identity may be measured.

"Human artificial chromosomes" (HACs) are linear microchromosomes which may contain DNA sequences of about 6 kb to 10 Mb in size and which contain all of the elements required for chromosome replication, segregation and maintenance.

The term "humanized antibody" refers to an antibody molecule in which the amino acid sequence in the non-antigen binding regions has been altered so that the antibody more closely resembles a human antibody, and still retains its original binding ability.

"Hybridization" refers to the process by which a polynucleotide strand anneals with a complementary strand through base pairing under defined hybridization conditions. Specific hybridization is an indication that two nucleic acid sequences share a high degree of complementarity. Specific hybridization complexes form under permissive annealing conditions and remain hybridized after the "washing" step(s). The washing step(s) is particularly important in determining the stringency of the hybridization process, with more stringent conditions allowing less non-specific binding, i.e., binding between pairs of nucleic acid strands that are not perfectly matched. Permissive conditions for annealing of nucleic acid sequences are routinely determinable by one of ordinary skill in the art and may be consistent among hybridization experiments, whereas wash conditions may be varied among experiments to achieve the desired stringency, and therefore hybridization specificity. Permissive annealing conditions occur, for example, at 68°C in the presence of about 6 x SSC, about 1% (w/v) SDS, and about 100 μg/ml sheared, denatured salmon sperm DNA.

Generally, stringency of hybridization is expressed, in part, with reference to the temperature under which the wash step is carried out. Such wash temperatures are typically selected to be about 5°C to 20°C lower than the thermal melting point (T_m) for the specific sequence at a defined ionic strength and pH. The T_m is the temperature (under defined ionic strength and pH) at which 50% of the target sequence hybridizes to a perfectly matched probe. An equation for calculating T_m and conditions for nucleic acid hybridization are well known and can be found in Sambrook, J. et al. (1989) <u>Molecular Cloning: A Laboratory Manual</u>, 2nd ed., vol. 1-3, Cold Spring Harbor Press, Plainview NY; specifically see volume 2, chapter 9.

High stringency conditions for hybridization between polynucleotides of the present invention include wash conditions of 68°C in the presence of about 0.2 x SSC and about 0.1% SDS, for 1 hour. Alternatively, temperatures of about 65°C, 60°C, 55°C, or 42°C may be used. SSC

concentration may be varied from about 0.1 to 2 x SSC, with SDS being present at about 0.1%. Typically, blocking reagents are used to block non-specific hybridization. Such blocking reagents include, for instance, sheared and denatured salmon sperm DNA at about 100-200 μg/ml. Organic solvent, such as formamide at a concentration of about 35-50% v/v, may also be used under particular circumstances, such as for RNA:DNA hybridizations. Useful variations on these wash conditions will be readily apparent to those of ordinary skill in the art. Hybridization, particularly under high stringency conditions, may be suggestive of evolutionary similarity between the nucleotides. Such similarity is strongly indicative of a similar role for the nucleotides and their encoded polypeptides.

The term "hybridization complex" refers to a complex formed between two nucleic acid sequences by virtue of the formation of hydrogen bonds between complementary bases. A hybridization complex may be formed in solution (e.g., C_0t or R_0t analysis) or formed between one nucleic acid sequence present in solution and another nucleic acid sequence immobilized on a solid support (e.g., paper, membranes, filters, chips, pins or glass slides, or any other appropriate substrate to which cells or their nucleic acids have been fixed).

The words "insertion" and "addition" refer to changes in an amino acid or nucleotide sequence resulting in the addition of one or more amino acid residues or nucleotides, respectively.

"Immune response" can refer to conditions associated with inflammation, trauma, immune disorders, or infectious or genetic disease, etc. These conditions can be characterized by expression of various factors, e.g., cytokines, chemokines, and other signaling molecules, which may affect cellular and systemic defense systems.

An "immunogenic fragment" is a polypeptide or oligopeptide fragment of TRICH which is capable of eliciting an immune response when introduced into a living organism, for example, a mammal. The term "immunogenic fragment" also includes any polypeptide or oligopeptide fragment of TRICH which is useful in any of the antibody production methods disclosed herein or known in the art.

The term "microarray" refers to an arrangement of a plurality of polynucleotides, polypeptides, or other chemical compounds on a substrate.

The terms "element" and "array element" refer to a polynucleotide, polypeptide, or other chemical compound having a unique and defined position on a microarray.

The term "modulate" refers to a change in the activity of TRICH. For example, modulation may cause an increase or a decrease in protein activity, binding characteristics, or any other biological, functional, or immunological properties of TRICH.

The phrases "nucleic acid" and "nucleic acid sequence" refer to a nucleotide, oligonucleotide, polynucleotide, or any fragment thereof. These phrases also refer to DNA or RNA of genomic or synthetic origin which may be single-stranded or double-stranded and may represent the sense or the

antisense strand, to peptide nucleic acid (PNA), or to any DNA-like or RNA-like material.

"Operably linked" refers to the situation in which a first nucleic acid sequence is placed in a functional relationship with a second nucleic acid sequence. For instance, a promoter is operably linked to a coding sequence if the promoter affects the transcription or expression of the coding sequence. Operably linked DNA sequences may be in close proximity or contiguous and, where necessary to join two protein coding regions, in the same reading frame.

"Peptide nucleic acid" (PNA) refers to an antisense molecule or anti-gene agent which comprises an oligonucleotide of at least about 5 nucleotides in length linked to a peptide backbone of amino acid residues ending in lysine. The terminal lysine confers solubility to the composition. PNAs preferentially bind complementary single stranded DNA or RNA and stop transcript elongation, and may be pegylated to extend their lifespan in the cell.

"Post-translational modification" of an TRICH may involve lipidation, glycosylation, phosphorylation, acetylation, racemization, proteolytic cleavage, and other modifications known in the art. These processes may occur synthetically or biochemically. Biochemical modifications will vary by cell type depending on the enzymatic milieu of TRICH.

"Probe" refers to nucleic acid sequences encoding TRICH, their complements, or fragments thereof, which are used to detect identical, allelic or related nucleic acid sequences. Probes are isolated oligonucleotides or polynucleotides attached to a detectable label or reporter molecule. Typical labels include radioactive isotopes, ligands, chemiluminescent agents, and enzymes. "Primers" are short nucleic acids, usually DNA oligonucleotides, which may be annealed to a target polynucleotide by complementary base-pairing. The primer may then be extended along the target DNA strand by a DNA polymerase enzyme. Primer pairs can be used for amplification (and identification) of a nucleic acid sequence, e.g., by the polymerase chain reaction (PCR).

Probes and primers as used in the present invention typically comprise at least 15 contiguous nucleotides of a known sequence. In order to enhance specificity, longer probes and primers may also be employed, such as probes and primers that comprise at least 20, 25, 30, 40, 50, 60, 70, 80, 90, 100, or at least 150 consecutive nucleotides of the disclosed nucleic acid sequences. Probes and primers may be considerably longer than these examples, and it is understood that any length supported by the specification, including the tables, figures, and Sequence Listing, may be used.

Methods for preparing and using probes and primers are described in the references, for example Sambrook, J. et al. (1989) Molecular Cloning: A Laboratory Manual, 2nd ed., vol. 1-3, Cold Spring Harbor Press, Plainview NY; Ausubel, F.M. et al. (1987) Current Protocols in Molecular Biology, Greene Publ. Assoc. & Wiley-Intersciences, New York NY; Innis, M. et al. (1990) PCR Protocols, A Guide to Methods and Applications, Academic Press, San Diego CA. PCR primer pairs can be derived from a known sequence, for example, by using computer programs intended for that

purpose such as Primer (Version 0.5, 1991, Whitehead Institute for Biomedical Research, Cambridge MA).

Oligonucleotides for use as primers are selected using software known in the art for such purpose. For example, OLIGO 4.06 software is useful for the selection of PCR primer pairs of up to 100 nucleotides each, and for the analysis of oligonucleotides and larger polynucleotides of up to 5,000 nucleotides from an input polynucleotide sequence of up to 32 kilobases. Similar primer selection programs have incorporated additional features for expanded capabilities. For example, the PrimOU primer selection program (available to the public from the Genome Center at University of Texas South West Medical Center, Dallas TX) is capable of choosing specific primers from megabase sequences and is thus useful for designing primers on a genome-wide scope. The Primer3 primer selection program (available to the public from the Whitehead Institute/MIT Center for Genome Research, Cambridge MA) allows the user to input a "mispriming library," in which sequences to avoid as primer binding sites are user-specified. Primer3 is useful, in particular, for the selection of oligonucleotides for microarrays. (The source code for the latter two primer selection programs may also be obtained from their respective sources and modified to meet the user's specific needs.) The PrimeGen program (available to the public from the UK Human Genome Mapping Project Resource Centre, Cambridge UK) designs primers based on multiple sequence alignments, thereby allowing selection of primers that hybridize to either the most conserved or least conserved regions of aligned nucleic acid sequences. Hence, this program is useful for identification of both unique and conserved oligonucleotides and polynucleotide fragments. The oligonucleotides and polynucleotide fragments identified by any of the above selection methods are useful in hybridization technologies, for example, as PCR or sequencing primers, microarray elements, or specific probes to identify fully or partially complementary polynucleotides in a sample of nucleic acids. Methods of oligonucleotide selection are not limited to those described above.

A "recombinant nucleic acid" is a sequence that is not naturally occurring or has a sequence that is made by an artificial combination of two or more otherwise separated segments of sequence. This artificial combination is often accomplished by chemical synthesis or, more commonly, by the artificial manipulation of isolated segments of nucleic acids, e.g., by genetic engineering techniques such as those described in Sambrook, supra. The term recombinant includes nucleic acids that have been altered solely by addition, substitution, or deletion of a portion of the nucleic acid. Frequently, a recombinant nucleic acid may include a nucleic acid sequence operably linked to a promoter sequence. Such a recombinant nucleic acid may be part of a vector that is used, for example, to transform a cell.

Alternatively, such recombinant nucleic acids may be part of a viral vector, e.g., based on a vaccinia virus, that could be use to vaccinate a mammal wherein the recombinant nucleic acid is

expressed, inducing a protective immunological response in the mammal.

A "regulatory element" refers to a nucleic acid sequence usually derived from untranslated regions of a gene and includes enhancers, promoters, introns, and 5' and 3' untranslated regions (UTRs). Regulatory elements interact with host or viral proteins which control transcription, translation, or RNA stability.

"Reporter molecules" are chemical or biochemical moieties used for labeling a nucleic acid, amino acid, or antibody. Reporter molecules include radionuclides; enzymes; fluorescent, chemiluminescent, or chromogenic agents; substrates; cofactors; inhibitors; magnetic particles; and other moieties known in the art.

An "RNA equivalent," in reference to a DNA sequence, is composed of the same linear sequence of nucleotides as the reference DNA sequence with the exception that all occurrences of the nitrogenous base thymine are replaced with uracil, and the sugar backbone is composed of ribose instead of deoxyribose.

The term "sample" is used in its broadest sense. A sample suspected of containing TRICH, nucleic acids encoding TRICH, or fragments thereof may comprise a bodily fluid; an extract from a cell, chromosome, organelle, or membrane isolated from a cell; a cell; genomic DNA, RNA, or cDNA, in solution or bound to a substrate; a tissue; a tissue print; etc.

The terms "specific binding" and "specifically binding" refer to that interaction between a protein or peptide and an agonist, an antibody, an antagonist, a small molecule, or any natural or synthetic binding composition. The interaction is dependent upon the presence of a particular structure of the protein, e.g., the antigenic determinant or epitope, recognized by the binding molecule. For example, if an antibody is specific for epitope "A," the presence of a polypeptide comprising the epitope A, or the presence of free unlabeled A, in a reaction containing free labeled A and the antibody will reduce the amount of labeled A that binds to the antibody.

The term "substantially purified" refers to nucleic acid or amino acid sequences that are removed from their natural environment and are isolated or separated, and are at least 60% free, preferably at least 75% free, and most preferably at least 90% free from other components with which they are naturally associated.

A "substitution" refers to the replacement of one or more amino acid residues or nucleotides by different amino acid residues or nucleotides, respectively.

"Substrate" refers to any suitable rigid or semi-rigid support including membranes, filters, chips, slides, wafers, fibers, magnetic or nonmagnetic beads, gels, tubing, plates, polymers, microparticles and capillaries. The substrate can have a variety of surface forms, such as wells, trenches, pins, channels and pores, to which polynucleotides or polypeptides are bound.

A "transcript image" refers to the collective pattern of gene expression by a particular cell

type or tissue under given conditions at a given time.

"Transformation" describes a process by which exogenous DNA is introduced into a recipient cell. Transformation may occur under natural or artificial conditions according to various methods well known in the art, and may rely on any known method for the insertion of foreign nucleic acid sequences into a prokaryotic or eukaryotic host cell. The method for transformation is selected based on the type of host cell being transformed and may include, but is not limited to, bacteriophage or viral infection, electroporation, heat shock, lipofection, and particle bombardment. The term "transformed cells" includes stably transformed cells in which the inserted DNA is capable of replication either as an autonomously replicating plasmid or as part of the host chromosome, as well as transiently transformed cells which express the inserted DNA or RNA for limited periods of time.

A "transgenic organism," as used herein, is any organism, including but not limited to animals and plants, in which one or more of the cells of the organism contains heterologous nucleic acid introduced by way of human intervention, such as by transgenic techniques well known in the art. The nucleic acid is introduced into the cell, directly or indirectly by introduction into a precursor of the cell, by way of deliberate genetic manipulation, such as by microinjection or by infection with a recombinant virus. The term genetic manipulation does not include classical cross-breeding, or in vitro fertilization, but rather is directed to the introduction of a recombinant DNA molecule. The transgenic organisms contemplated in accordance with the present invention include bacteria, cyanobacteria, fungi, plants and animals. The isolated DNA of the present invention can be introduced into the host by methods known in the art, for example infection, transfection, transformation or transconjugation. Techniques for transferring the DNA of the present invention into such organisms are widely known and provided in references such as Sambrook et al. (1989), supra.

A "variant" of a particular nucleic acid sequence is defined as a nucleic acid sequence having at least 40% sequence identity to the particular nucleic acid sequence over a certain length of one of the nucleic acid sequences using blastn with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) set at default parameters. Such a pair of nucleic acids may show, for example, at least 50%, at least 60%, at least 70%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% or greater sequence identity over a certain defined length. A variant may be described as, for example, an "allelic" (as defined above), "splice," "species," or "polymorphic" variant. A splice variant may have significant identity to a reference molecule, but will generally have a greater or lesser number of polynucleotides due to alternate splicing of exons during mRNA processing. The corresponding polypeptide may possess additional functional domains or lack domains that are present in the reference molecule. Species variants are polynucleotide sequences that vary from one species to

another. The resulting polypeptides will generally have significant amino acid identity relative to each other. A polymorphic variant is a variation in the polynucleotide sequence of a particular gene between individuals of a given species. Polymorphic variants also may encompass "single nucleotide polymorphisms" (SNPs) in which the polynucleotide sequence varies by one nucleotide base. The presence of SNPs may be indicative of, for example, a certain population, a disease state, or a propensity for a disease state.

A "variant" of a particular polypeptide sequence is defined as a polypeptide sequence having at least 40% sequence identity to the particular polypeptide sequence over a certain length of one of the polypeptide sequences using blastp with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) set at default parameters. Such a pair of polypeptides may show, for example, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% or greater sequence identity over a certain defined length of one of the polypeptides.

THE INVENTION

The invention is based on the discovery of new human transporters and ion channels (TRICH), the polynucleotides encoding TRICH, and the use of these compositions for the diagnosis, treatment, or prevention of transport, neurological, muscle, immunological, and cell proliferative disorders.

Table 1 summarizes the nomenclature for the full length polynucleotide and polypeptide sequences of the invention. Each polynucleotide and its corresponding polypeptide are correlated to a single Incyte project identification number (Incyte Project ID). Each polypeptide sequence is denoted by both a polypeptide sequence identification number (Polypeptide SEQ ID NO:) and an Incyte polypeptide sequence number (Incyte Polypeptide ID) as shown. Each polynucleotide sequence is denoted by both a polynucleotide sequence identification number (Polynucleotide SEQ ID NO:) and an Incyte polynucleotide consensus sequence number (Incyte Polynucleotide ID) as shown.

Table 2 shows sequences with homology to the polypeptides of the invention as identified by BLAST analysis against the GenBank protein (genpept) database. Columns 1 and 2 show the polypeptide sequence identification number (Polypeptide SEQ ID NO:) and the corresponding Incyte polypeptide sequence number (Incyte Polypeptide ID) for polypeptides of the invention. Column 3 shows the GenBank identification number (Genbank ID NO:) of the nearest GenBank homolog. Column 4 shows the probability score for the match between each polypeptide and its GenBank homolog. Column 5 shows the annotation of the GenBank homolog along with relevant citations where applicable, all of which are expressly incorporated by reference herein.

Table 3 shows various structural features of the polypeptides of the invention. Columns 1

and 2 show the polypeptide sequence identification number (SEQ ID NO:) and the corresponding Incyte polypeptide sequence number (Incyte Polypeptide ID) for each polypeptide of the invention. Column 3 shows the number of amino acid residues in each polypeptide. Column 4 shows potential phosphorylation sites, and column 5 shows potential glycosylation sites, as determined by the MOTIFS program of the GCG sequence analysis software package (Genetics Computer Group, Madison WI). Column 6 shows amino acid residues comprising signature sequences, domains, and motifs. Column 7 shows analytical methods for protein structure/function analysis and in some cases, searchable databases to which the analytical methods were applied.

Together, Tables 2 and 3 summarize the properties of polypeptides of the invention, and these properties establish that the claimed polypeptides are transporters and ion channels. For example, SEQ ID NO:6 is 89% identical to rat neuronal nicotinic acetylcholine receptor subunit (GenBank ID g6746563) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 1.7e-188, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:6 also contains a neurotransmitter-gated ion channel domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from BLIMPS, MOTIFS, and PROFILESCAN analyses provide further corroborative evidence that SEQ ID NO:6 is a neurotransmitter-gated ion channel. In an alternative example, SEQ ID NO:14 is 93% identical to rat TAP-like ABC transporter (GenBank ID g6045150) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 0.0, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:14 also contains an ABC transporter domain and an ABC transporter transmembrane region as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from BLIMPS, MOTIFS, and PROFILESCAN analyses provide further corroborative evidence that SEQ ID NO:14 is an ABC transporter. In an alternative example, SEQ ID NO:16 is 98% identical to human voltage-dependent anion channel (GenBank ID g340199) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 1.2e-130, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:16 also contains a eukaryotic porin active site domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from BLIMPS, MOTIFS, and PROFILESCAN analyses provide further corroborative evidence that SEQ ID NO:16 is a mitochondrial porin. In an alternative example, SEQ ID NO:20 is 28% identical to a rat voltage-gated calcium channel (GenBank ID g4586963) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The

BLAST probability score is 2.4e-27, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. Data from BLIMPS and BLAST analyses provide further corroborative evidence that SEQ ID NO:20 is a voltage-gated calcium channel. In an alternative example, SEQ ID NO:22 is 82% identical to human inhibitory glycine receptor (GenBank ID g31849) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 1.1e-175, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:22 also contains a neurotransmitter-gated ion channel domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from BLIMPS, MOTIFS, and PROFILESCAN analyses provide further corroborative evidence that SEQ ID NO:22 is a neurotransmitter-gated ion channel. In an alternative example, SEQ ID NO:30 is 36% identical to human ATP binding cassette (ABC) -C transporter (GenBank ID g1514530) as determined by the Basic Local Alignment Search Tool (BLAST, see Table 2). The BLAST probability score is 2.3e-127, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:30 also contains ABC transporter domains as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains (see Table 3). Data from BLIMPS, MOTIFS, and PROFILESCAN analyses provide further corroborative evidence that SEQ ID NO:30 is an ABC transporter. SEQ ID NO:1-5, SEQ ID NO:7-13, SEQ ID NO:15, SEQ ID NO:17-19, SEQ ID NO:21, and SEQ ID NO:23-29 were analyzed and annotated in a similar manner. The algorithms and parameters for the analysis of SEQ ID NO:1-30 are described in Table 7.

As shown in Table 4, the full length polynucleotide sequences of the present invention were assembled using cDNA sequences or coding (exon) sequences derived from genomic DNA, or any combination of these two types of sequences. Columns 1 and 2 list the polynucleotide sequence identification number (Polynucleotide SEQ ID NO:) and the corresponding Incyte polynucleotide consensus sequence number (Incyte Polynucleotide ID) for each polynucleotide of the invention. Column 3 shows the length of each polynucleotide sequence in basepairs. Column 4 lists fragments of the polynucleotide sequences which are useful, for example, in hybridization or amplification technologies that identify SEQ ID NO:31-60 or that distinguish between SEQ ID NO:31-60 and related polynucleotide sequences. Column 5 shows identification numbers corresponding to cDNA sequences, coding sequences (exons) predicted from genomic DNA, and/or sequence assemblages comprised of both cDNA and genomic DNA. These sequences were used to assemble the full length polynucleotide sequences of the invention. Columns 6 and 7 of Table 4 show the nucleotide start (5') and stop (3') positions of the cDNA and/or genomic sequences in column 5 relative to their respective full length sequences.

The identification numbers in Column 5 of Table 4 may refer specifically, for example, to Incyte cDNAs along with their corresponding cDNA libraries. For example, 6340750H1 is the identification number of an Incyte cDNA sequence, and BRANDIN01 is the cDNA library from which it is derived. Incyte cDNAs for which cDNA libraries are not indicated were derived from pooled cDNA libraries (e.g., 71911330V1). Alternatively, the identification numbers in column 5 may refer to GenBank cDNAs or ESTs (e.g., g5110579) which contributed to the assembly of the full length polynucleotide sequences. In addition, the identification numbers in column 5 may identify sequences derived from the ENSEMBL (The Sanger Centre, Cambridge, UK) database (*i.e.*, those sequences including the designation "ENST"). Alternatively, the identification numbers in column 5 may be derived from the NCBI RefSeq Nucleotide Sequence Records Database (*i.e.*, those sequences including the designation "NM" or "NT") or the NCBI RefSeq Protein Sequence Records (*i.e.*, those sequences including the designation "NP"). Alternatively, the identification numbers in column 5 may refer to assemblages of both cDNA and Genscan-predicted exons brought together by an "exon stitching" algorithm. For example, FL_$XXXXXX_N_1_N_2_YYYYY_N_3_N_4$ represents a "stitched" sequence in which *XXXXXX* is the identification number of the cluster of sequences to which the algorithm was applied, and *YYYYY* is the number of the prediction generated by the algorithm, and $N_{1,2,3...}$, if present, represent specific exons that may have been manually edited during analysis (See Example V). Alternatively, the identification numbers in column 5 may refer to assemblages of exons brought together by an "exon-stretching" algorithm. For example, FL*XXXXXX*_*gAAAAA*_*gBBBBB*_1_*N* is the identification number of a "stretched" sequence, with *XXXXXX* being the Incyte project identification number, *gAAAAA* being the GenBank identification number of the human genomic sequence to which the "exon-stretching" algorithm was applied, *gBBBBB* being the GenBank identification number or NCBI RefSeq identification number of the nearest GenBank protein homolog, and *N* referring to specific exons (See Example V). In instances where a RefSeq sequence was used as a protein homolog for the "exon-stretching" algorithm, a RefSeq identifier (denoted by "NM," "NP," or "NT") may be used in place of the GenBank identifier (*i.e.*, *gBBBBB*).

Alternatively, a prefix identifies component sequences that were hand-edited, predicted from genomic DNA sequences, or derived from a combination of sequence analysis methods. The following Table lists examples of component sequence prefixes and corresponding sequence analysis methods associated with the prefixes (see Example IV and Example V).

Prefix	Type of analysis and/or examples of programs

GNN, GFG, ENST	Exon prediction from genomic sequences using, for example, GENSCAN (Stanford University, CA, USA) or FGENES (Computer Genomics Group, The Sanger Centre, Cambridge, UK).
GBI	Hand-edited analysis of genomic sequences.
FL	Stitched or stretched genomic sequences (see Example V).
INCY	Full length transcript and exon prediction from mapping of EST sequences to the genome. Genomic location and EST composition data are combined to predict the exons and resulting transcript.

In some cases, Incyte cDNA coverage redundant with the sequence coverage shown in column 5 was obtained to confirm the final consensus polynucleotide sequence, but the relevant Incyte cDNA identification numbers are not shown.

Table 5 shows the representative cDNA libraries for those full length polynucleotide sequences which were assembled using Incyte cDNA sequences. The representative cDNA library is the Incyte cDNA library which is most frequently represented by the Incyte cDNA sequences which were used to assemble and confirm the above polynucleotide sequences. The tissues and vectors which were used to construct the cDNA libraries shown in Table 5 are described in Table 6.

The invention also encompasses TRICH variants. A preferred TRICH variant is one which has at least about 80%, or alternatively at least about 90%, or even at least about 95% amino acid sequence identity to the TRICH amino acid sequence, and which contains at least one functional or structural characteristic of TRICH.

The invention also encompasses polynucleotides which encode TRICH. In a particular embodiment, the invention encompasses a polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:31-60, which encodes TRICH. The polynucleotide sequences of SEQ ID NO:31-60, as presented in the Sequence Listing, embrace the equivalent RNA sequences, wherein occurrences of the nitrogenous base thymine are replaced with uracil, and the sugar backbone is composed of ribose instead of deoxyribose.

The invention also encompasses a variant of a polynucleotide sequence encoding TRICH. In particular, such a variant polynucleotide sequence will have at least about 70%, or alternatively at least about 85%, or even at least about 95% polynucleotide sequence identity to the polynucleotide sequence encoding TRICH. A particular aspect of the invention encompasses a variant of a polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:31-60 which has at least about 70%, or alternatively at least about 85%, or even at least about 95% polynucleotide sequence identity to a nucleic acid sequence selected from the group consisting

of SEQ ID NO:31-60. Any one of the polynucleotide variants described above can encode an amino acid sequence which contains at least one functional or structural characteristic of TRICH.

It will be appreciated by those skilled in the art that as a result of the degeneracy of the genetic code, a multitude of polynucleotide sequences encoding TRICH, some bearing minimal similarity to the polynucleotide sequences of any known and naturally occurring gene, may be produced. Thus, the invention contemplates each and every possible variation of polynucleotide sequence that could be made by selecting combinations based on possible codon choices. These combinations are made in accordance with the standard triplet genetic code as applied to the polynucleotide sequence of naturally occurring TRICH, and all such variations are to be considered as being specifically disclosed.

Although nucleotide sequences which encode TRICH and its variants are generally capable of hybridizing to the nucleotide sequence of the naturally occurring TRICH under appropriately selected conditions of stringency, it may be advantageous to produce nucleotide sequences encoding TRICH or its derivatives possessing a substantially different codon usage, e.g., inclusion of non-naturally occurring codons. Codons may be selected to increase the rate at which expression of the peptide occurs in a particular prokaryotic or eukaryotic host in accordance with the frequency with which particular codons are utilized by the host. Other reasons for substantially altering the nucleotide sequence encoding TRICH and its derivatives without altering the encoded amino acid sequences include the production of RNA transcripts having more desirable properties, such as a greater half-life, than transcripts produced from the naturally occurring sequence.

The invention also encompasses production of DNA sequences which encode TRICH and TRICH derivatives, or fragments thereof, entirely by synthetic chemistry. After production, the synthetic sequence may be inserted into any of the many available expression vectors and cell systems using reagents well known in the art. Moreover, synthetic chemistry may be used to introduce mutations into a sequence encoding TRICH or any fragment thereof.

Also encompassed by the invention are polynucleotide sequences that are capable of hybridizing to the claimed polynucleotide sequences, and, in particular, to those shown in SEQ ID NO:31-60 and fragments thereof under various conditions of stringency. (See, e.g., Wahl, G.M. and S.L. Berger (1987) Methods Enzymol. 152:399-407; Kimmel, A.R. (1987) Methods Enzymol. 152:507-511.) Hybridization conditions, including annealing and wash conditions, are described in "Definitions."

Methods for DNA sequencing are well known in the art and may be used to practice any of the embodiments of the invention. The methods may employ such enzymes as the Klenow fragment of DNA polymerase I, SEQUENASE (US Biochemical, Cleveland OH), Taq polymerase (Applied Biosystems), thermostable T7 polymerase (Amersham Pharmacia Biotech, Piscataway NJ), or

combinations of polymerases and proofreading exonucleases such as those found in the ELONGASE amplification system (Life Technologies, Gaithersburg MD). Preferably, sequence preparation is automated with machines such as the MICROLAB 2200 liquid transfer system (Hamilton, Reno NV), PTC200 thermal cycler (MJ Research, Watertown MA) and ABI CATALYST 800 thermal cycler (Applied Biosystems). Sequencing is then carried out using either the ABI 373 or 377 DNA sequencing system (Applied Biosystems), the MEGABACE 1000 DNA sequencing system (Molecular Dynamics, Sunnyvale CA), or other systems known in the art. The resulting sequences are analyzed using a variety of algorithms which are well known in the art. (See, e.g., Ausubel, F.M. (1997) Short Protocols in Molecular Biology, John Wiley & Sons, New York NY, unit 7.7; Meyers, R.A. (1995) Molecular Biology and Biotechnology, Wiley VCH, New York NY, pp. 856-853.)

The nucleic acid sequences encoding TRICH may be extended utilizing a partial nucleotide sequence and employing various PCR-based methods known in the art to detect upstream sequences, such as promoters and regulatory elements. For example, one method which may be employed, restriction-site PCR, uses universal and nested primers to amplify unknown sequence from genomic DNA within a cloning vector. (See, e.g., Sarkar, G. (1993) PCR Methods Applic. 2:318-322.) Another method, inverse PCR, uses primers that extend in divergent directions to amplify unknown sequence from a circularized template. The template is derived from restriction fragments comprising a known genomic locus and surrounding sequences. (See, e.g., Triglia, T. et al. (1988) Nucleic Acids Res. 16:8186.) A third method, capture PCR, involves PCR amplification of DNA fragments adjacent to known sequences in human and yeast artificial chromosome DNA. (See, e.g., Lagerstrom, M. et al. (1991) PCR Methods Applic. 1:111-119.) In this method, multiple restriction enzyme digestions and ligations may be used to insert an engineered double-stranded sequence into a region of unknown sequence before performing PCR. Other methods which may be used to retrieve unknown sequences are known in the art. (See, e.g., Parker, J.D. et al. (1991) Nucleic Acids Res. 19:3055-3060). Additionally, one may use PCR, nested primers, and PROMOTERFINDER libraries (Clontech, Palo Alto CA) to walk genomic DNA. This procedure avoids the need to screen libraries and is useful in finding intron/exon junctions. For all PCR-based methods, primers may be designed using commercially available software, such as OLIGO 4.06 primer analysis software (National Biosciences, Plymouth MN) or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the template at temperatures of about 68°C to 72°C.

When screening for full length cDNAs, it is preferable to use libraries that have been size-selected to include larger cDNAs. In addition, random-primed libraries, which often include sequences containing the 5' regions of genes, are preferable for situations in which an oligo d(T) library does not yield a full-length cDNA. Genomic libraries may be useful for extension of sequence

into 5' non-transcribed regulatory regions.

Capillary electrophoresis systems which are commercially available may be used to analyze the size or confirm the nucleotide sequence of sequencing or PCR products. In particular, capillary sequencing may employ flowable polymers for electrophoretic separation, four different nucleotide-specific, laser-stimulated fluorescent dyes, and a charge coupled device camera for detection of the emitted wavelengths. Output/light intensity may be converted to electrical signal using appropriate software (e.g., GENOTYPER and SEQUENCE NAVIGATOR, Applied Biosystems), and the entire process from loading of samples to computer analysis and electronic data display may be computer controlled. Capillary electrophoresis is especially preferable for sequencing small DNA fragments which may be present in limited amounts in a particular sample.

In another embodiment of the invention, polynucleotide sequences or fragments thereof which encode TRICH may be cloned in recombinant DNA molecules that direct expression of TRICH, or fragments or functional equivalents thereof, in appropriate host cells. Due to the inherent degeneracy of the genetic code, other DNA sequences which encode substantially the same or a functionally equivalent amino acid sequence may be produced and used to express TRICH.

The nucleotide sequences of the present invention can be engineered using methods generally known in the art in order to alter TRICH-encoding sequences for a variety of purposes including, but not limited to, modification of the cloning, processing, and/or expression of the gene product. DNA shuffling by random fragmentation and PCR reassembly of gene fragments and synthetic oligonucleotides may be used to engineer the nucleotide sequences. For example, oligonucleotide-mediated site-directed mutagenesis may be used to introduce mutations that create new restriction sites, alter glycosylation patterns, change codon preference, produce splice variants, and so forth.

The nucleotides of the present invention may be subjected to DNA shuffling techniques such as MOLECULARBREEDING (Maxygen Inc., Santa Clara CA; described in U.S. Patent Number 5,837,458; Chang, C.-C. et al. (1999) Nat. Biotechnol. 17:793-797; Christians, F.C. et al. (1999) Nat. Biotechnol. 17:259-264; and Crameri, A. et al. (1996) Nat. Biotechnol. 14:315-319) to alter or improve the biological properties of TRICH, such as its biological or enzymatic activity or its ability to bind to other molecules or compounds. DNA shuffling is a process by which a library of gene variants is produced using PCR-mediated recombination of gene fragments. The library is then subjected to selection or screening procedures that identify those gene variants with the desired properties. These preferred variants may then be pooled and further subjected to recursive rounds of DNA shuffling and selection/screening. Thus, genetic diversity is created through "artificial" breeding and rapid molecular evolution. For example, fragments of a single gene containing random point mutations may be recombined, screened, and then reshuffled until the desired properties are optimized. Alternatively, fragments of a given gene may be recombined with fragments of

homologous genes in the same gene family, either from the same or different species, thereby maximizing the genetic diversity of multiple naturally occurring genes in a directed and controllable manner.

In another embodiment, sequences encoding TRICH may be synthesized, in whole or in part, using chemical methods well known in the art. (See, e.g., Caruthers, M.H. et al. (1980) Nucleic Acids Symp. Ser. 7:215-223; and Horn, T. et al. (1980) Nucleic Acids Symp. Ser. 7:225-232.) Alternatively, TRICH itself or a fragment thereof may be synthesized using chemical methods. For example, peptide synthesis can be performed using various solution-phase or solid-phase techniques. (See, e.g., Creighton, T. (1984) Proteins, Structures and Molecular Properties, WH Freeman, New York NY, pp. 55-60; and Roberge, J.Y. et al. (1995) Science 269:202-204.) Automated synthesis may be achieved using the ABI 431A peptide synthesizer (Applied Biosystems). Additionally, the amino acid sequence of TRICH, or any part thereof, may be altered during direct synthesis and/or combined with sequences from other proteins, or any part thereof, to produce a variant polypeptide or a polypeptide having a sequence of a naturally occurring polypeptide.

The peptide may be substantially purified by preparative high performance liquid chromatography. (See, e.g., Chiez, R.M. and F.Z. Regnier (1990) Methods Enzymol. 182:392-421.) The composition of the synthetic peptides may be confirmed by amino acid analysis or by sequencing. (See, e.g., Creighton, supra, pp. 28-53.)

In order to express a biologically active TRICH, the nucleotide sequences encoding TRICH or derivatives thereof may be inserted into an appropriate expression vector, i.e., a vector which contains the necessary elements for transcriptional and translational control of the inserted coding sequence in a suitable host. These elements include regulatory sequences, such as enhancers, constitutive and inducible promoters, and 5' and 3' untranslated regions in the vector and in polynucleotide sequences encoding TRICH. Such elements may vary in their strength and specificity. Specific initiation signals may also be used to achieve more efficient translation of sequences encoding TRICH. Such signals include the ATG initiation codon and adjacent sequences, e.g. the Kozak sequence. In cases where sequences encoding TRICH and its initiation codon and upstream regulatory sequences are inserted into the appropriate expression vector, no additional transcriptional or translational control signals may be needed. However, in cases where only coding sequence, or a fragment thereof, is inserted, exogenous translational control signals including an in-frame ATG initiation codon should be provided by the vector. Exogenous translational elements and initiation codons may be of various origins, both natural and synthetic. The efficiency of expression may be enhanced by the inclusion of enhancers appropriate for the particular host cell system used. (See, e.g., Scharf, D. et al. (1994) Results Probl. Cell Differ. 20:125-162.)

Methods which are well known to those skilled in the art may be used to construct expression

vectors containing sequences encoding TRICH and appropriate transcriptional and translational control elements. These methods include in vitro recombinant DNA techniques, synthetic techniques, and in vivo genetic recombination. (See, e.g., Sambrook, J. et al. (1989) Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Press, Plainview NY, ch. 4, 8, and 16-17; Ausubel, F.M. et al. (1995) Current Protocols in Molecular Biology, John Wiley & Sons, New York NY, ch. 9, 13, and 16.)

A variety of expression vector/host systems may be utilized to contain and express sequences encoding TRICH. These include, but are not limited to, microorganisms such as bacteria transformed with recombinant bacteriophage, plasmid, or cosmid DNA expression vectors; yeast transformed with yeast expression vectors; insect cell systems infected with viral expression vectors (e.g., baculovirus); plant cell systems transformed with viral expression vectors (e.g., cauliflower mosaic virus, CaMV, or tobacco mosaic virus, TMV) or with bacterial expression vectors (e.g., Ti or pBR322 plasmids); or animal cell systems. (See, e.g., Sambrook, supra; Ausubel, supra; Van Heeke, G. and S.M. Schuster (1989) J. Biol. Chem. 264:5503-5509; Engelhard, E.K. et al. (1994) Proc. Natl. Acad. Sci. USA 91:3224-3227; Sandig, V. et al. (1996) Hum. Gene Ther. 7:1937-1945; Takamatsu, N. (1987) EMBO J. 6:307-311; The McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York NY, pp. 191-196; Logan, J. and T. Shenk (1984) Proc. Natl. Acad. Sci. USA 81:3655-3659; and Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355.) Expression vectors derived from retroviruses, adenoviruses, or herpes or vaccinia viruses, or from various bacterial plasmids, may be used for delivery of nucleotide sequences to the targeted organ, tissue, or cell population. (See, e.g., Di Nicola, M. et al. (1998) Cancer Gen. Ther. 5(6):350-356; Yu, M. et al. (1993) Proc. Natl. Acad. Sci. USA 90(13):6340-6344; Buller, R.M. et al. (1985) Nature 317(6040):813-815; McGregor, D.P. et al. (1994) Mol. Immunol. 31(3):219-226; and Verma, I.M. and N. Somia (1997) Nature 389:239-242.) The invention is not limited by the host cell employed.

In bacterial systems, a number of cloning and expression vectors may be selected depending upon the use intended for polynucleotide sequences encoding TRICH. For example, routine cloning, subcloning, and propagation of polynucleotide sequences encoding TRICH can be achieved using a multifunctional E. coli vector such as PBLUESCRIPT (Stratagene, La Jolla CA) or PSPORT1 plasmid (Life Technologies). Ligation of sequences encoding TRICH into the vector's multiple cloning site disrupts the *lacZ* gene, allowing a colorimetric screening procedure for identification of transformed bacteria containing recombinant molecules. In addition, these vectors may be useful for in vitro transcription, dideoxy sequencing, single strand rescue with helper phage, and creation of nested deletions in the cloned sequence. (See, e.g., Van Heeke, G. and S.M. Schuster (1989) J. Biol. Chem. 264:5503-5509.) When large quantities of TRICH are needed, e.g. for the production of antibodies, vectors which direct high level expression of TRICH may be used. For example, vectors

containing the strong, inducible SP6 or T7 bacteriophage promoter may be used.

Yeast expression systems may be used for production of TRICH. A number of vectors containing constitutive or inducible promoters, such as alpha factor, alcohol oxidase, and PGH promoters, may be used in the yeast Saccharomyces cerevisiae or Pichia pastoris. In addition, such vectors direct either the secretion or intracellular retention of expressed proteins and enable integration of foreign sequences into the host genome for stable propagation. (See, e.g., Ausubel, 1995, supra; Bitter, G.A. et al. (1987) Methods Enzymol. 153:516-544; and Scorer, C.A. et al. (1994) Bio/Technology 12:181-184.)

Plant systems may also be used for expression of TRICH. Transcription of sequences encoding TRICH may be driven by viral promoters, e.g., the 35S and 19S promoters of CaMV used alone or in combination with the omega leader sequence from TMV (Takamatsu, N. (1987) EMBO J. 6:307-311). Alternatively, plant promoters such as the small subunit of RUBISCO or heat shock promoters may be used. (See, e.g., Coruzzi, G. et al. (1984) EMBO J. 3:1671-1680; Broglie, R. et al. (1984) Science 224:838-843; and Winter, J. et al. (1991) Results Probl. Cell Differ. 17:85-105.) These constructs can be introduced into plant cells by direct DNA transformation or pathogen-mediated transfection. (See, e.g., The McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York NY, pp. 191-196.)

In mammalian cells, a number of viral-based expression systems may be utilized. In cases where an adenovirus is used as an expression vector, sequences encoding TRICH may be ligated into an adenovirus transcription/translation complex consisting of the late promoter and tripartite leader sequence. Insertion in a non-essential E1 or E3 region of the viral genome may be used to obtain infective virus which expresses TRICH in host cells. (See, e.g., Logan, J. and T. Shenk (1984) Proc. Natl. Acad. Sci. USA 81:3655-3659.) In addition, transcription enhancers, such as the Rous sarcoma virus (RSV) enhancer, may be used to increase expression in mammalian host cells. SV40 or EBV-based vectors may also be used for high-level protein expression.

Human artificial chromosomes (HACs) may also be employed to deliver larger fragments of DNA than can be contained in and expressed from a plasmid. HACs of about 6 kb to 10 Mb are constructed and delivered via conventional delivery methods (liposomes, polycationic amino polymers, or vesicles) for therapeutic purposes. (See, e.g., Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355.)

For long term production of recombinant proteins in mammalian systems, stable expression of TRICH in cell lines is preferred. For example, sequences encoding TRICH can be transformed into cell lines using expression vectors which may contain viral origins of replication and/or endogenous expression elements and a selectable marker gene on the same or on a separate vector. Following the introduction of the vector, cells may be allowed to grow for about 1 to 2 days in

enriched media before being switched to selective media. The purpose of the selectable marker is to confer resistance to a selective agent, and its presence allows growth and recovery of cells which successfully express the introduced sequences. Resistant clones of stably transformed cells may be propagated using tissue culture techniques appropriate to the cell type.

Any number of selection systems may be used to recover transformed cell lines. These include, but are not limited to, the herpes simplex virus thymidine kinase and adenine phosphoribosyltransferase genes, for use in *tk⁻* and *apr⁻* cells, respectively. (See, e.g., Wigler, M. et al. (1977) Cell 11:223-232; Lowy, I. et al. (1980) Cell 22:817-823.) Also, antimetabolite, antibiotic, or herbicide resistance can be used as the basis for selection. For example, *dhfr* confers resistance to methotrexate; *neo* confers resistance to the aminoglycosides neomycin and G-418; and *als* and *pat* confer resistance to chlorsulfuron and phosphinotricin acetyltransferase, respectively. (See, e.g., Wigler, M. et al. (1980) Proc. Natl. Acad. Sci. USA 77:3567-3570; Colbere-Garapin, F. et al. (1981) J. Mol. Biol. 150:1-14.) Additional selectable genes have been described, e.g., *trpB* and *hisD*, which alter cellular requirements for metabolites. (See, e.g., Hartman, S.C. and R.C. Mulligan (1988) Proc. Natl. Acad. Sci. USA 85:8047-8051.) Visible markers, e.g., anthocyanins, green fluorescent proteins (GFP; Clontech), ß glucuronidase and its substrate ß-glucuronide, or luciferase and its substrate luciferin may be used. These markers can be used not only to identify transformants, but also to quantify the amount of transient or stable protein expression attributable to a specific vector system. (See, e.g., Rhodes, C.A. (1995) Methods Mol. Biol. 55:121-131.)

Although the presence/absence of marker gene expression suggests that the gene of interest is also present, the presence and expression of the gene may need to be confirmed. For example, if the sequence encoding TRICH is inserted within a marker gene sequence, transformed cells containing sequences encoding TRICH can be identified by the absence of marker gene function. Alternatively, a marker gene can be placed in tandem with a sequence encoding TRICH under the control of a single promoter. Expression of the marker gene in response to induction or selection usually indicates expression of the tandem gene as well.

In general, host cells that contain the nucleic acid sequence encoding TRICH and that express TRICH may be identified by a variety of procedures known to those of skill in the art. These procedures include, but are not limited to, DNA-DNA or DNA-RNA hybridizations, PCR amplification, and protein bioassay or immunoassay techniques which include membrane, solution, or chip based technologies for the detection and/or quantification of nucleic acid or protein sequences.

Immunological methods for detecting and measuring the expression of TRICH using either specific polyclonal or monoclonal antibodies are known in the art. Examples of such techniques include enzyme-linked immunosorbent assays (ELISAs), radioimmunoassays (RIAs), and fluorescence activated cell sorting (FACS). A two-site, monoclonal-based immunoassay utilizing

monoclonal antibodies reactive to two non-interfering epitopes on TRICH is preferred, but a competitive binding assay may be employed. These and other assays are well known in the art. (See, e.g., Hampton, R. et al. (1990) Serological Methods, a Laboratory Manual, APS Press, St. Paul MN, Sect. IV; Coligan, J.E. et al. (1997) Current Protocols in Immunology, Greene Pub. Associates and Wiley-Interscience, New York NY; and Pound, J.D. (1998) Immunochemical Protocols, Humana Press, Totowa NJ.)

A wide variety of labels and conjugation techniques are known by those skilled in the art and may be used in various nucleic acid and amino acid assays. Means for producing labeled hybridization or PCR probes for detecting sequences related to polynucleotides encoding TRICH include oligolabeling, nick translation, end-labeling, or PCR amplification using a labeled nucleotide. Alternatively, the sequences encoding TRICH, or any fragments thereof, may be cloned into a vector for the production of an mRNA probe. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes in vitro by addition of an appropriate RNA polymerase such as T7, T3, or SP6 and labeled nucleotides. These procedures may be conducted using a variety of commercially available kits, such as those provided by Amersham Pharmacia Biotech, Promega (Madison WI), and US Biochemical. Suitable reporter molecules or labels which may be used for ease of detection include radionuclides, enzymes, fluorescent, chemiluminescent, or chromogenic agents, as well as substrates, cofactors, inhibitors, magnetic particles, and the like.

Host cells transformed with nucleotide sequences encoding TRICH may be cultured under conditions suitable for the expression and recovery of the protein from cell culture. The protein produced by a transformed cell may be secreted or retained intracellularly depending on the sequence and/or the vector used. As will be understood by those of skill in the art, expression vectors containing polynucleotides which encode TRICH may be designed to contain signal sequences which direct secretion of TRICH through a prokaryotic or eukaryotic cell membrane.

In addition, a host cell strain may be chosen for its ability to modulate expression of the inserted sequences or to process the expressed protein in the desired fashion. Such modifications of the polypeptide include, but are not limited to, acetylation, carboxylation, glycosylation, phosphorylation, lipidation, and acylation. Post-translational processing which cleaves a "prepro" or "pro" form of the protein may also be used to specify protein targeting, folding, and/or activity. Different host cells which have specific cellular machinery and characteristic mechanisms for post-translational activities (e.g., CHO, HeLa, MDCK, HEK293, and WI38) are available from the American Type Culture Collection (ATCC, Manassas VA) and may be chosen to ensure the correct modification and processing of the foreign protein.

In another embodiment of the invention, natural, modified, or recombinant nucleic acid sequences encoding TRICH may be ligated to a heterologous sequence resulting in translation of a

fusion protein in any of the aforementioned host systems. For example, a chimeric TRICH protein containing a heterologous moiety that can be recognized by a commercially available antibody may facilitate the screening of peptide libraries for inhibitors of TRICH activity. Heterologous protein and peptide moieties may also facilitate purification of fusion proteins using commercially available affinity matrices. Such moieties include, but are not limited to, glutathione S-transferase (GST), maltose binding protein (MBP), thioredoxin (Trx), calmodulin binding peptide (CBP), 6-His, FLAG, *c-myc*, and hemagglutinin (HA). GST, MBP, Trx, CBP, and 6-His enable purification of their cognate fusion proteins on immobilized glutathione, maltose, phenylarsine oxide, calmodulin, and metal-chelate resins, respectively. FLAG, *c-myc*, and hemagglutinin (HA) enable immunoaffinity purification of fusion proteins using commercially available monoclonal and polyclonal antibodies that specifically recognize these epitope tags. A fusion protein may also be engineered to contain a proteolytic cleavage site located between the TRICH encoding sequence and the heterologous protein sequence, so that TRICH may be cleaved away from the heterologous moiety following purification. Methods for fusion protein expression and purification are discussed in Ausubel (1995, supra, ch. 10). A variety of commercially available kits may also be used to facilitate expression and purification of fusion proteins.

In a further embodiment of the invention, synthesis of radiolabeled TRICH may be achieved in vitro using the TNT rabbit reticulocyte lysate or wheat germ extract system (Promega). These systems couple transcription and translation of protein-coding sequences operably associated with the T7, T3, or SP6 promoters. Translation takes place in the presence of a radiolabeled amino acid precursor, for example, ^{35}S-methionine.

TRICH of the present invention or fragments thereof may be used to screen for compounds that specifically bind to TRICH. At least one and up to a plurality of test compounds may be screened for specific binding to TRICH. Examples of test compounds include antibodies, oligonucleotides, proteins (e.g., receptors), or small molecules.

In one embodiment, the compound thus identified is closely related to the natural ligand of TRICH, e.g., a ligand or fragment thereof, a natural substrate, a structural or functional mimetic, or a natural binding partner. (See, e.g., Coligan, J.E. et al. (1991) Current Protocols in Immunology 1(2): Chapter 5.) Similarly, the compound can be closely related to the natural receptor to which TRICH binds, or to at least a fragment of the receptor, e.g., the ligand binding site. In either case, the compound can be rationally designed using known techniques. In one embodiment, screening for these compounds involves producing appropriate cells which express TRICH, either as a secreted protein or on the cell membrane. Preferred cells include cells from mammals, yeast, Drosophila, or E. coli. Cells expressing TRICH or cell membrane fractions which contain TRICH are then contacted with a test compound and binding, stimulation, or inhibition of activity of either TRICH or the

compound is analyzed.

An assay may simply test binding of a test compound to the polypeptide, wherein binding is detected by a fluorophore, radioisotope, enzyme conjugate, or other detectable label. For example, the assay may comprise the steps of combining at least one test compound with TRICH, either in solution or affixed to a solid support, and detecting the binding of TRICH to the compound. Alternatively, the assay may detect or measure binding of a test compound in the presence of a labeled competitor. Additionally, the assay may be carried out using cell-free preparations, chemical libraries, or natural product mixtures, and the test compound(s) may be free in solution or affixed to a solid support.

TRICH of the present invention or fragments thereof may be used to screen for compounds that modulate the activity of TRICH. Such compounds may include agonists, antagonists, or partial or inverse agonists. In one embodiment, an assay is performed under conditions permissive for TRICH activity, wherein TRICH is combined with at least one test compound, and the activity of TRICH in the presence of a test compound is compared with the activity of TRICH in the absence of the test compound. A change in the activity of TRICH in the presence of the test compound is indicative of a compound that modulates the activity of TRICH. Alternatively, a test compound is combined with an in vitro or cell-free system comprising TRICH under conditions suitable for TRICH activity, and the assay is performed. In either of these assays, a test compound which modulates the activity of TRICH may do so indirectly and need not come in direct contact with the test compound. At least one and up to a plurality of test compounds may be screened.

In another embodiment, polynucleotides encoding TRICH or their mammalian homologs may be "knocked out" in an animal model system using homologous recombination in embryonic stem (ES) cells. Such techniques are well known in the art and are useful for the generation of animal models of human disease. (See, e.g., U.S. Patent Number 5,175,383 and U.S. Patent Number 5,767,337.) For example, mouse ES cells, such as the mouse 129/SvJ cell line, are derived from the early mouse embryo and grown in culture. The ES cells are transformed with a vector containing the gene of interest disrupted by a marker gene, e.g., the neomycin phosphotransferase gene (neo; Capecchi, M.R. (1989) Science 244:1288-1292). The vector integrates into the corresponding region of the host genome by homologous recombination. Alternatively, homologous recombination takes place using the Cre-loxP system to knockout a gene of interest in a tissue- or developmental stage-specific manner (Marth, J.D. (1996) Clin. Invest. 97:1999-2002; Wagner, K.U. et al. (1997) Nucleic Acids Res. 25:4323-4330). Transformed ES cells are identified and microinjected into mouse cell blastocysts such as those from the C57BL/6 mouse strain. The blastocysts are surgically transferred to pseudopregnant dams, and the resulting chimeric progeny are genotyped and bred to produce heterozygous or homozygous strains. Transgenic animals thus generated may be tested with potential

therapeutic or toxic agents.

Polynucleotides encoding TRICH may also be manipulated in vitro in ES cells derived from human blastocysts. Human ES cells have the potential to differentiate into at least eight separate cell lineages including endoderm, mesoderm, and ectodermal cell types. These cell lineages differentiate into, for example, neural cells, hematopoietic lineages, and cardiomyocytes (Thomson, J.A. et al. (1998) Science 282:1145-1147).

Polynucleotides encoding TRICH can also be used to create "knockin" humanized animals (pigs) or transgenic animals (mice or rats) to model human disease. With knockin technology, a region of a polynucleotide encoding TRICH is injected into animal ES cells, and the injected sequence integrates into the animal cell genome. Transformed cells are injected into blastulae, and the blastulae are implanted as described above. Transgenic progeny or inbred lines are studied and treated with potential pharmaceutical agents to obtain information on treatment of a human disease. Alternatively, a mammal inbred to overexpress TRICH, e.g., by secreting TRICH in its milk, may also serve as a convenient source of that protein (Janne, J. et al. (1998) Biotechnol. Annu. Rev. 4:55-74).

THERAPEUTICS

Chemical and structural similarity, e.g., in the context of sequences and motifs, exists between regions of TRICH and transporters and ion channels. In addition, the expression of TRICH is closely associated with brain, liver, tumor, colon, thymus, small intestine, myometrium, testicular, bone marrow neuroblastoma tumor, parotid gland, lung, pituitary gland, and placental tissues, and Pompe's disease. Therefore, TRICH appears to play a role in transport, neurological, muscle, immunological, and cell proliferative disorders. In the treatment of disorders associated with increased TRICH expression or activity, it is desirable to decrease the expression or activity of TRICH. In the treatment of disorders associated with decreased TRICH expression or activity, it is desirable to increase the expression or activity of TRICH.

Therefore, in one embodiment, TRICH or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of TRICH. Examples of such disorders include, but are not limited to, a transport disorder such as akinesia, amyotrophic lateral sclerosis, ataxia telangiectasia, cystic fibrosis, Becker's muscular dystrophy, Bell's palsy, Charcot-Marie Tooth disease, diabetes mellitus, diabetes insipidus, diabetic neuropathy, Duchenne muscular dystrophy, hyperkalemic periodic paralysis, normokalemic periodic paralysis, Parkinson's disease, malignant hyperthermia, multidrug resistance, myasthenia gravis, myotonic dystrophy, catatonia, tardive dyskinesia, dystonias, peripheral neuropathy, cerebral neoplasms, prostate cancer, cardiac disorders associated with transport, e.g., angina, bradyarrythmia, tachyarrythmia, hypertension, Long QT syndrome, myocarditis, cardiomyopathy, nemaline myopathy, centronuclear myopathy, lipid myopathy, mitochondrial myopathy, thyrotoxic myopathy,

ethanol myopathy, dermatomyositis, inclusion body myositis, infectious myositis, polymyositis, neurological disorders associated with transport, e.g., Alzheimer's disease, amnesia, bipolar disorder, dementia, depression, epilepsy, Tourette's disorder, paranoid psychoses, and schizophrenia, and other disorders associated with transport, e.g., neurofibromatosis, postherpetic neuralgia, trigeminal neuropathy, sarcoidosis, sickle cell anemia, Wilson's disease, cataracts, infertility, pulmonary artery stenosis, sensorineural autosomal deafness, hyperglycemia, hypoglycemia, Grave's disease, goiter, Cushing's disease, Addison's disease, glucose-galactose malabsorption syndrome, hypercholesterolemia, adrenoleukodystrophy, Zellweger syndrome, Menkes disease, occipital horn syndrome, von Gierke disease, cystinuria, iminoglycinuria, Hartup disease, and Fanconi disease; a neurological disorder such as epilepsy, ischemic cerebrovascular disease, stroke, cerebral neoplasms, Alzheimer's disease, Pick's disease, Huntington's disease, dementia, Parkinson's disease and other extrapyramidal disorders, amyotrophic lateral sclerosis and other motor neuron disorders, progressive neural muscular atrophy, retinitis pigmentosa, hereditary ataxias, multiple sclerosis and other demyelinating diseases, bacterial and viral meningitis, brain abscess, subdural empyema, epidural abscess, suppurative intracranial thrombophlebitis, myelitis and radiculitis, viral central nervous system disease, prion diseases including kuru, Creutzfeldt-Jakob disease, and Gerstmann-Straussler-Scheinker syndrome, fatal familial insomnia, nutritional and metabolic diseases of the nervous system, neurofibromatosis, tuberous sclerosis, cerebelloretinal hemangioblastomatosis, encephalotrigeminal syndrome, mental retardation and other developmental disorders of the central nervous system including Down syndrome, cerebral palsy, neuroskeletal disorders, autonomic nervous system disorders, cranial nerve disorders, spinal cord diseases, muscular dystrophy and other neuromuscular disorders, peripheral nervous system disorders, dermatomyositis and polymyositis, inherited, metabolic, endocrine, and toxic myopathies, myasthenia gravis, periodic paralysis, mental disorders including mood, anxiety, and schizophrenic disorders, seasonal affective disorder (SAD), akathesia, amnesia, catatonia, diabetic neuropathy, tardive dyskinesia, dystonias, paranoid psychoses, postherpetic neuralgia, Tourette's disorder, progressive supranuclear palsy, corticobasal degeneration, and familial frontotemporal dementia; a muscle disorder such as cardiomyopathy, myocarditis, Duchenne's muscular dystrophy, Becker's muscular dystrophy, myotonic dystrophy, central core disease, nemaline myopathy, centronuclear myopathy, lipid myopathy, mitochondrial myopathy, infectious myositis, polymyositis, dermatomyositis, inclusion body myositis, thyrotoxic myopathy, ethanol myopathy, angina, anaphylactic shock, arrhythmias, asthma, cardiovascular shock, Cushing's syndrome, hypertension, hypoglycemia, myocardial infarction, migraine, pheochromocytoma, and myopathies including encephalopathy, epilepsy, Kearns-Sayre syndrome, lactic acidosis, myoclonic disorder, ophthalmoplegia, and acid maltase deficiency (AMD, also known

as Pompe's disease); an immunological disorder such as acquired immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED), bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, and trauma; and a cell proliferative disorder such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and cancers including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus.

In another embodiment, a vector capable of expressing TRICH or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of TRICH including, but not limited to, those described above.

In a further embodiment, a composition comprising a substantially purified TRICH in conjunction with a suitable pharmaceutical carrier may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of TRICH including, but not limited to, those provided above.

In still another embodiment, an agonist which modulates the activity of TRICH may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of TRICH including, but not limited to, those listed above.

In a further embodiment, an antagonist of TRICH may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of TRICH. Examples of such disorders include, but are not limited to, those transport, neurological, muscle, immunological, and cell proliferative disorders described above. In one aspect, an antibody which specifically binds

TRICH may be used directly as an antagonist or indirectly as a targeting or delivery mechanism for bringing a pharmaceutical agent to cells or tissues which express TRICH.

In an additional embodiment, a vector expressing the complement of the polynucleotide encoding TRICH may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of TRICH including, but not limited to, those described above.

In other embodiments, any of the proteins, antagonists, antibodies, agonists, complementary sequences, or vectors of the invention may be administered in combination with other appropriate therapeutic agents. Selection of the appropriate agents for use in combination therapy may be made by one of ordinary skill in the art, according to conventional pharmaceutical principles. The combination of therapeutic agents may act synergistically to effect the treatment or prevention of the various disorders described above. Using this approach, one may be able to achieve therapeutic efficacy with lower dosages of each agent, thus reducing the potential for adverse side effects.

An antagonist of TRICH may be produced using methods which are generally known in the art. In particular, purified TRICH may be used to produce antibodies or to screen libraries of pharmaceutical agents to identify those which specifically bind TRICH. Antibodies to TRICH may also be generated using methods that are well known in the art. Such antibodies may include, but are not limited to, polyclonal, monoclonal, chimeric, and single chain antibodies, Fab fragments, and fragments produced by a Fab expression library. Neutralizing antibodies (i.e., those which inhibit dimer formation) are generally preferred for therapeutic use.

For the production of antibodies, various hosts including goats, rabbits, rats, mice, humans, and others may be immunized by injection with TRICH or with any fragment or oligopeptide thereof which has immunogenic properties. Depending on the host species, various adjuvants may be used to increase immunological response. Such adjuvants include, but are not limited to, Freund's, mineral gels such as aluminum hydroxide, and surface active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, KLH, and dinitrophenol. Among adjuvants used in humans, BCG (bacilli Calmette-Guerin) and Corynebacterium parvum are especially preferable.

It is preferred that the oligopeptides, peptides, or fragments used to induce antibodies to TRICH have an amino acid sequence consisting of at least about 5 amino acids, and generally will consist of at least about 10 amino acids. It is also preferable that these oligopeptides, peptides, or fragments are identical to a portion of the amino acid sequence of the natural protein. Short stretches of TRICH amino acids may be fused with those of another protein, such as KLH, and antibodies to the chimeric molecule may be produced.

Monoclonal antibodies to TRICH may be prepared using any technique which provides for the production of antibody molecules by continuous cell lines in culture. These include, but are not limited to, the hybridoma technique, the human B-cell hybridoma technique, and the EBV-hybridoma

technique. (See, e.g., Kohler, G. et al. (1975) Nature 256:495-497; Kozbor, D. et al. (1985) J. Immunol. Methods 81:31-42; Cote, R.J. et al. (1983) Proc. Natl. Acad. Sci. USA 80:2026-2030; and Cole, S.P. et al. (1984) Mol. Cell Biol. 62:109-120.)

In addition, techniques developed for the production of "chimeric antibodies," such as the splicing of mouse antibody genes to human antibody genes to obtain a molecule with appropriate antigen specificity and biological activity, can be used. (See, e.g., Morrison, S.L. et al. (1984) Proc. Natl. Acad. Sci. USA 81:6851-6855; Neuberger, M.S. et al. (1984) Nature 312:604-608; and Takeda, S. et al. (1985) Nature 314:452-454.) Alternatively, techniques described for the production of single chain antibodies may be adapted, using methods known in the art, to produce TRICH-specific single chain antibodies. Antibodies with related specificity, but of distinct idiotypic composition, may be generated by chain shuffling from random combinatorial immunoglobulin libraries. (See, e.g., Burton, D.R. (1991) Proc. Natl. Acad. Sci. USA 88:10134-10137.)

Antibodies may also be produced by inducing in vivo production in the lymphocyte population or by screening immunoglobulin libraries or panels of highly specific binding reagents as disclosed in the literature. (See, e.g., Orlandi, R. et al. (1989) Proc. Natl. Acad. Sci. USA 86:3833-3837; Winter, G. et al. (1991) Nature 349:293-299.)

Antibody fragments which contain specific binding sites for TRICH may also be generated. For example, such fragments include, but are not limited to, $F(ab')_2$ fragments produced by pepsin digestion of the antibody molecule and Fab fragments generated by reducing the disulfide bridges of the F(ab')2 fragments. Alternatively, Fab expression libraries may be constructed to allow rapid and easy identification of monoclonal Fab fragments with the desired specificity. (See, e.g., Huse, W.D. et al. (1989) Science 246:1275-1281.)

Various immunoassays may be used for screening to identify antibodies having the desired specificity. Numerous protocols for competitive binding or immunoradiometric assays using either polyclonal or monoclonal antibodies with established specificities are well known in the art. Such immunoassays typically involve the measurement of complex formation between TRICH and its specific antibody. A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering TRICH epitopes is generally used, but a competitive binding assay may also be employed (Pound, supra).

Various methods such as Scatchard analysis in conjunction with radioimmunoassay techniques may be used to assess the affinity of antibodies for TRICH. Affinity is expressed as an association constant, K_a, which is defined as the molar concentration of TRICH-antibody complex divided by the molar concentrations of free antigen and free antibody under equilibrium conditions. The K_a determined for a preparation of polyclonal antibodies, which are heterogeneous in their affinities for multiple TRICH epitopes, represents the average affinity, or avidity, of the antibodies

for TRICH. The K_a determined for a preparation of monoclonal antibodies, which are monospecific for a particular TRICH epitope, represents a true measure of affinity. High-affinity antibody preparations with K_a ranging from about 10^9 to 10^{12} L/mole are preferred for use in immunoassays in which the TRICH-antibody complex must withstand rigorous manipulations. Low-affinity antibody preparations with K_a ranging from about 10^6 to 10^7 L/mole are preferred for use in immunopurification and similar procedures which ultimately require dissociation of TRICH, preferably in active form, from the antibody (Catty, D. (1988) Antibodies, Volume I: A Practical Approach, IRL Press, Washington DC; Liddell, J.E. and A. Cryer (1991) A Practical Guide to Monoclonal Antibodies, John Wiley & Sons, New York NY).

The titer and avidity of polyclonal antibody preparations may be further evaluated to determine the quality and suitability of such preparations for certain downstream applications. For example, a polyclonal antibody preparation containing at least 1-2 mg specific antibody/ml, preferably 5-10 mg specific antibody/ml, is generally employed in procedures requiring precipitation of TRICH-antibody complexes. Procedures for evaluating antibody specificity, titer, and avidity, and guidelines for antibody quality and usage in various applications, are generally available. (See, e.g., Catty, supra, and Coligan et al. supra.)

In another embodiment of the invention, the polynucleotides encoding TRICH, or any fragment or complement thereof, may be used for therapeutic purposes. In one aspect, modifications of gene expression can be achieved by designing complementary sequences or antisense molecules (DNA, RNA, PNA, or modified oligonucleotides) to the coding or regulatory regions of the gene encoding TRICH. Such technology is well known in the art, and antisense oligonucleotides or larger fragments can be designed from various locations along the coding or control regions of sequences encoding TRICH. (See, e.g., Agrawal, S., ed. (1996) Antisense Therapeutics, Humana Press Inc., Totawa NJ.)

In therapeutic use, any gene delivery system suitable for introduction of the antisense sequences into appropriate target cells can be used. Antisense sequences can be delivered intracellularly in the form of an expression plasmid which, upon transcription, produces a sequence complementary to at least a portion of the cellular sequence encoding the target protein. (See, e.g., Slater, J.E. et al. (1998) J. Allergy Clin. Immunol. 102(3):469-475; and Scanlon, K.J. et al. (1995) 9(13):1288-1296.) Antisense sequences can also be introduced intracellularly through the use of viral vectors, such as retrovirus and adeno-associated virus vectors. (See, e.g., Miller, A.D. (1990) Blood 76:271; Ausubel, supra; Uckert, W. and W. Walther (1994) Pharmacol. Ther. 63(3):323-347.) Other gene delivery mechanisms include liposome-derived systems, artificial viral envelopes, and other systems known in the art. (See, e.g., Rossi, J.J. (1995) Br. Med. Bull. 51(1):217-225; Boado, R.J. et al. (1998) J. Pharm. Sci. 87(11):1308-1315; and Morris, M.C. et al. (1997) Nucleic Acids Res.

25(14):2730-2736.)

In another embodiment of the invention, polynucleotides encoding TRICH may be used for somatic or germline gene therapy. Gene therapy may be performed to (i) correct a genetic deficiency (e.g., in the cases of severe combined immunodeficiency (SCID)-X1 disease characterized by X-linked inheritance (Cavazzana-Calvo, M. et al. (2000) Science 288:669-672), severe combined immunodeficiency syndrome associated with an inherited adenosine deaminase (ADA) deficiency (Blaese, R.M. et al. (1995) Science 270:475-480; Bordignon, C. et al. (1995) Science 270:470-475), cystic fibrosis (Zabner, J. et al. (1993) Cell 75:207-216; Crystal, R.G. et al. (1995) Hum. Gene Therapy 6:643-666; Crystal, R.G. et al. (1995) Hum. Gene Therapy 6:667-703), thalassamias, familial hypercholesterolemia, and hemophilia resulting from Factor VIII or Factor IX deficiencies (Crystal, R.G. (1995) Science 270:404-410; Verma, I.M. and N. Somia (1997) Nature 389:239-242)), (ii) express a conditionally lethal gene product (e.g., in the case of cancers which result from unregulated cell proliferation), or (iii) express a protein which affords protection against intracellular parasites (e.g., against human retroviruses, such as human immunodeficiency virus (HIV) (Baltimore, D. (1988) Nature 335:395-396; Poeschla, E. et al. (1996) Proc. Natl. Acad. Sci. USA. 93:11395-11399), hepatitis B or C virus (HBV, HCV); fungal parasites, such as Candida albicans and Paracoccidioides brasiliensis; and protozoan parasites such as Plasmodium falciparum and Trypanosoma cruzi). In the case where a genetic deficiency in TRICH expression or regulation causes disease, the expression of TRICH from an appropriate population of transduced cells may alleviate the clinical manifestations caused by the genetic deficiency.

In a further embodiment of the invention, diseases or disorders caused by deficiencies in TRICH are treated by constructing mammalian expression vectors encoding TRICH and introducing these vectors by mechanical means into TRICH-deficient cells. Mechanical transfer technologies for use with cells in vivo or ex vitro include (i) direct DNA microinjection into individual cells, (ii) ballistic gold particle delivery, (iii) liposome-mediated transfection, (iv) receptor-mediated gene transfer, and (v) the use of DNA transposons (Morgan, R.A. and W.F. Anderson (1993) Annu. Rev. Biochem. 62:191-217; Ivics, Z. (1997) Cell 91:501-510; Boulay, J-L. and H. Récipon (1998) Curr. Opin. Biotechnol. 9:445-450).

Expression vectors that may be effective for the expression of TRICH include, but are not limited to, the PCDNA 3.1, EPITAG, PRCCMV2, PREP, PVAX, PCR2-TOPOTA vectors (Invitrogen, Carlsbad CA), PCMV-SCRIPT, PCMV-TAG, PEGSH/PERV (Stratagene, La Jolla CA), and PTET-OFF, PTET-ON, PTRE2, PTRE2-LUC, PTK-HYG (Clontech, Palo Alto CA). TRICH may be expressed using (i) a constitutively active promoter, (e.g., from cytomegalovirus (CMV), Rous sarcoma virus (RSV), SV40 virus, thymidine kinase (TK), or β-actin genes), (ii) an inducible promoter (e.g., the tetracycline-regulated promoter (Gossen, M. and H. Bujard (1992) Proc. Natl.

Acad. Sci. USA 89:5547-5551; Gossen, M. et al. (1995) Science 268:1766-1769; Rossi, F.M.V. and H.M. Blau (1998) Curr. Opin. Biotechnol. 9:451-456), commercially available in the T-REX plasmid (Invitrogen)); the ecdysone-inducible promoter (available in the plasmids PVGRXR and PIND; Invitrogen); the FK506/rapamycin inducible promoter; or the RU486/mifepristone inducible promoter (Rossi, F.M.V. and Blau, H.M. supra)), or (iii) a tissue-specific promoter or the native promoter of the endogenous gene encoding TRICH from a normal individual.

Commercially available liposome transformation kits (e.g., the PERFECT LIPID TRANSFECTION KIT, available from Invitrogen) allow one with ordinary skill in the art to deliver polynucleotides to target cells in culture and require minimal effort to optimize experimental parameters. In the alternative, transformation is performed using the calcium phosphate method (Graham, F.L. and A.J. Eb (1973) Virology 52:456-467), or by electroporation (Neumann, E. et al. (1982) EMBO J. 1:841-845). The introduction of DNA to primary cells requires modification of these standardized mammalian transfection protocols.

In another embodiment of the invention, diseases or disorders caused by genetic defects with respect to TRICH expression are treated by constructing a retrovirus vector consisting of (i) the polynucleotide encoding TRICH under the control of an independent promoter or the retrovirus long terminal repeat (LTR) promoter, (ii) appropriate RNA packaging signals, and (iii) a Rev-responsive element (RRE) along with additional retrovirus *cis*-acting RNA sequences and coding sequences required for efficient vector propagation. Retrovirus vectors (e.g., PFB and PFBNEO) are commercially available (Stratagene) and are based on published data (Riviere, I. et al. (1995) Proc. Natl. Acad. Sci. USA 92:6733-6737), incorporated by reference herein. The vector is propagated in an appropriate vector producing cell line (VPCL) that expresses an envelope gene with a tropism for receptors on the target cells or a promiscuous envelope protein such as VSVg (Armentano, D. et al. (1987) J. Virol. 61:1647-1650; Bender, M.A. et al. (1987) J. Virol. 61:1639-1646; Adam, M.A. and A.D. Miller (1988) J. Virol. 62:3802-3806; Dull, T. et al. (1998) J. Virol. 72:8463-8471; Zufferey, R. et al. (1998) J. Virol. 72:9873-9880). U.S. Patent Number 5,910,434 to Rigg ("Method for obtaining retrovirus packaging cell lines producing high transducing efficiency retroviral supernatant") discloses a method for obtaining retrovirus packaging cell lines and is hereby incorporated by reference. Propagation of retrovirus vectors, transduction of a population of cells (e.g., $CD4^+$ T-cells), and the return of transduced cells to a patient are procedures well known to persons skilled in the art of gene therapy and have been well documented (Ranga, U. et al. (1997) J. Virol. 71:7020-7029; Bauer, G. et al. (1997) Blood 89:2259-2267; Bonyhadi, M.L. (1997) J. Virol. 71:4707-4716; Ranga, U. et al. (1998) Proc. Natl. Acad. Sci. USA 95:1201-1206; Su, L. (1997) Blood 89:2283-2290).

In the alternative, an adenovirus-based gene therapy delivery system is used to deliver

polynucleotides encoding TRICH to cells which have one or more genetic abnormalities with respect to the expression of TRICH. The construction and packaging of adenovirus-based vectors are well known to those with ordinary skill in the art. Replication defective adenovirus vectors have proven to be versatile for importing genes encoding immunoregulatory proteins into intact islets in the pancreas (Csete, M.E. et al. (1995) Transplantation 27:263-268). Potentially useful adenoviral vectors are described in U.S. Patent Number 5,707,618 to Armentano ("Adenovirus vectors for gene therapy"), hereby incorporated by reference. For adenoviral vectors, see also Antinozzi, P.A. et al. (1999) Annu. Rev. Nutr. 19:511-544 and Verma, I.M. and N. Somia (1997) Nature 18:389:239-242, both incorporated by reference herein.

In another alternative, a herpes-based, gene therapy delivery system is used to deliver polynucleotides encoding TRICH to target cells which have one or more genetic abnormalities with respect to the expression of TRICH. The use of herpes simplex virus (HSV)-based vectors may be especially valuable for introducing TRICH to cells of the central nervous system, for which HSV has a tropism. The construction and packaging of herpes-based vectors are well known to those with ordinary skill in the art. A replication-competent herpes simplex virus (HSV) type 1-based vector has been used to deliver a reporter gene to the eyes of primates (Liu, X. et al. (1999) Exp. Eye Res. 169:385-395). The construction of a HSV-1 virus vector has also been disclosed in detail in U.S. Patent Number 5,804,413 to DeLuca ("Herpes simplex virus strains for gene transfer"), which is hereby incorporated by reference. U.S. Patent Number 5,804,413 teaches the use of recombinant HSV d92 which consists of a genome containing at least one exogenous gene to be transferred to a cell under the control of the appropriate promoter for purposes including human gene therapy. Also taught by this patent are the construction and use of recombinant HSV strains deleted for ICP4, ICP27 and ICP22. For HSV vectors, see also Goins, W.F. et al. (1999) J. Virol. 73:519-532 and Xu, H. et al. (1994) Dev. Biol. 163:152-161, hereby incorporated by reference. The manipulation of cloned herpesvirus sequences, the generation of recombinant virus following the transfection of multiple plasmids containing different segments of the large herpesvirus genomes, the growth and propagation of herpesvirus, and the infection of cells with herpesvirus are techniques well known to those of ordinary skill in the art.

In another alternative, an alphavirus (positive, single-stranded RNA virus) vector is used to deliver polynucleotides encoding TRICH to target cells. The biology of the prototypic alphavirus, Semliki Forest Virus (SFV), has been studied extensively and gene transfer vectors have been based on the SFV genome (Garoff, H. and K.-J. Li (1998) Curr. Opin. Biotechnol. 9:464-469). During alphavirus RNA replication, a subgenomic RNA is generated that normally encodes the viral capsid proteins. This subgenomic RNA replicates to higher levels than the full length genomic RNA, resulting in the overproduction of capsid proteins relative to the viral proteins with enzymatic activity

(e.g., protease and polymerase). Similarly, inserting the coding sequence for TRICH into the alphavirus genome in place of the capsid-coding region results in the production of a large number of TRICH-coding RNAs and the synthesis of high levels of TRICH in vector transduced cells. While alphavirus infection is typically associated with cell lysis within a few days, the ability to establish a persistent infection in hamster normal kidney cells (BHK-21) with a variant of Sindbis virus (SIN) indicates that the lytic replication of alphaviruses can be altered to suit the needs of the gene therapy application (Dryga, S.A. et al. (1997) Virology 228:74-83). The wide host range of alphaviruses will allow the introduction of TRICH into a variety of cell types. The specific transduction of a subset of cells in a population may require the sorting of cells prior to transduction. The methods of manipulating infectious cDNA clones of alphaviruses, performing alphavirus cDNA and RNA transfections, and performing alphavirus infections, are well known to those with ordinary skill in the art.

Oligonucleotides derived from the transcription initiation site, e.g., between about positions -10 and +10 from the start site, may also be employed to inhibit gene expression. Similarly, inhibition can be achieved using triple helix base-pairing methodology. Triple helix pairing is useful because it causes inhibition of the ability of the double helix to open sufficiently for the binding of polymerases, transcription factors, or regulatory molecules. Recent therapeutic advances using triplex DNA have been described in the literature. (See, e.g., Gee, J.E. et al. (1994) in Huber, B.E. and B.I. Carr, Molecular and Immunologic Approaches, Futura Publishing, Mt. Kisco NY, pp. 163-177.) A complementary sequence or antisense molecule may also be designed to block translation of mRNA by preventing the transcript from binding to ribosomes.

Ribozymes, enzymatic RNA molecules, may also be used to catalyze the specific cleavage of RNA. The mechanism of ribozyme action involves sequence-specific hybridization of the ribozyme molecule to complementary target RNA, followed by endonucleolytic cleavage. For example, engineered hammerhead motif ribozyme molecules may specifically and efficiently catalyze endonucleolytic cleavage of sequences encoding TRICH.

Specific ribozyme cleavage sites within any potential RNA target are initially identified by scanning the target molecule for ribozyme cleavage sites, including the following sequences: GUA, GUU, and GUC. Once identified, short RNA sequences of between 15 and 20 ribonucleotides, corresponding to the region of the target gene containing the cleavage site, may be evaluated for secondary structural features which may render the oligonucleotide inoperable. The suitability of candidate targets may also be evaluated by testing accessibility to hybridization with complementary oligonucleotides using ribonuclease protection assays.

Complementary ribonucleic acid molecules and ribozymes of the invention may be prepared by any method known in the art for the synthesis of nucleic acid molecules. These include techniques

for chemically synthesizing oligonucleotides such as solid phase phosphoramidite chemical synthesis. Alternatively, RNA molecules may be generated by in vitro and in vivo transcription of DNA sequences encoding TRICH. Such DNA sequences may be incorporated into a wide variety of vectors with suitable RNA polymerase promoters such as T7 or SP6. Alternatively, these cDNA constructs that synthesize complementary RNA, constitutively or inducibly, can be introduced into cell lines, cells, or tissues.

RNA molecules may be modified to increase intracellular stability and half-life. Possible modifications include, but are not limited to, the addition of flanking sequences at the 5' and/or 3' ends of the molecule, or the use of phosphorothioate or 2' O-methyl rather than phosphodiesterase linkages within the backbone of the molecule. This concept is inherent in the production of PNAs and can be extended in all of these molecules by the inclusion of nontraditional bases such as inosine, queosine, and wybutosine, as well as acetyl-, methyl-, thio-, and similarly modified forms of adenine, cytidine, guanine, thymine, and uridine which are not as easily recognized by endogenous endonucleases.

An additional embodiment of the invention encompasses a method for screening for a compound which is effective in altering expression of a polynucleotide encoding TRICH. Compounds which may be effective in altering expression of a specific polynucleotide may include, but are not limited to, oligonucleotides, antisense oligonucleotides, triple helix-forming oligonucleotides, transcription factors and other polypeptide transcriptional regulators, and non-macromolecular chemical entities which are capable of interacting with specific polynucleotide sequences. Effective compounds may alter polynucleotide expression by acting as either inhibitors or promoters of polynucleotide expression. Thus, in the treatment of disorders associated with increased TRICH expression or activity, a compound which specifically inhibits expression of the polynucleotide encoding TRICH may be therapeutically useful, and in the treatment of disorders associated with decreased TRICH expression or activity, a compound which specifically promotes expression of the polynucleotide encoding TRICH may be therapeutically useful.

At least one, and up to a plurality, of test compounds may be screened for effectiveness in altering expression of a specific polynucleotide. A test compound may be obtained by any method commonly known in the art, including chemical modification of a compound known to be effective in altering polynucleotide expression; selection from an existing, commercially-available or proprietary library of naturally-occurring or non-natural chemical compounds; rational design of a compound based on chemical and/or structural properties of the target polynucleotide; and selection from a library of chemical compounds created combinatorially or randomly. A sample comprising a polynucleotide encoding TRICH is exposed to at least one test compound thus obtained. The sample may comprise, for example, an intact or permeabilized cell, or an in vitro cell-free or reconstituted

biochemical system. Alterations in the expression of a polynucleotide encoding TRICH are assayed by any method commonly known in the art. Typically, the expression of a specific nucleotide is detected by hybridization with a probe having a nucleotide sequence complementary to the sequence of the polynucleotide encoding TRICH. The amount of hybridization may be quantified, thus forming the basis for a comparison of the expression of the polynucleotide both with and without exposure to one or more test compounds. Detection of a change in the expression of a polynucleotide exposed to a test compound indicates that the test compound is effective in altering the expression of the polynucleotide. A screen for a compound effective in altering expression of a specific polynucleotide can be carried out, for example, using a Schizosaccharomyces pombe gene expression system (Atkins, D. et al. (1999) U.S. Patent No. 5,932,435; Arndt, G.M. et al. (2000) Nucleic Acids Res. 28:E15) or a human cell line such as HeLa cell (Clarke, M.L. et al. (2000) Biochem. Biophys. Res. Commun. 268:8-13). A particular embodiment of the present invention involves screening a combinatorial library of oligonucleotides (such as deoxyribonucleotides, ribonucleotides, peptide nucleic acids, and modified oligonucleotides) for antisense activity against a specific polynucleotide sequence (Bruice, T.W. et al. (1997) U.S. Patent No. 5,686,242; Bruice, T.W. et al. (2000) U.S. Patent No. 6,022,691).

Many methods for introducing vectors into cells or tissues are available and equally suitable for use in vivo, in vitro, and ex vivo. For ex vivo therapy, vectors may be introduced into stem cells taken from the patient and clonally propagated for autologous transplant back into that same patient. Delivery by transfection, by liposome injections, or by polycationic amino polymers may be achieved using methods which are well known in the art. (See, e.g., Goldman, C.K. et al. (1997) Nat. Biotechnol. 15:462-466.)

Any of the therapeutic methods described above may be applied to any subject in need of such therapy, including, for example, mammals such as humans, dogs, cats, cows, horses, rabbits, and monkeys.

An additional embodiment of the invention relates to the administration of a composition which generally comprises an active ingredient formulated with a pharmaceutically acceptable excipient. Excipients may include, for example, sugars, starches, celluloses, gums, and proteins. Various formulations are commonly known and are thoroughly discussed in the latest edition of Remington's Pharmaceutical Sciences (Maack Publishing, Easton PA). Such compositions may consist of TRICH, antibodies to TRICH, and mimetics, agonists, antagonists, or inhibitors of TRICH.

The compositions utilized in this invention may be administered by any number of routes including, but not limited to, oral, intravenous, intramuscular, intra-arterial, intramedullary, intrathecal, intraventricular, pulmonary, transdermal, subcutaneous, intraperitoneal, intranasal, enteral, topical, sublingual, or rectal means.

Compositions for pulmonary administration may be prepared in liquid or dry powder form. These compositions are generally aerosolized immediately prior to inhalation by the patient. In the case of small molecules (e.g. traditional low molecular weight organic drugs), aerosol delivery of fast-acting formulations is well-known in the art. In the case of macromolecules (e.g. larger peptides and proteins), recent developments in the field of pulmonary delivery via the alveolar region of the lung have enabled the practical delivery of drugs such as insulin to blood circulation (see, e.g., Patton, J.S. et al., U.S. Patent No. 5,997,848). Pulmonary delivery has the advantage of administration without needle injection, and obviates the need for potentially toxic penetration enhancers.

Compositions suitable for use in the invention include compositions wherein the active ingredients are contained in an effective amount to achieve the intended purpose. The determination of an effective dose is well within the capability of those skilled in the art.

Specialized forms of compositions may be prepared for direct intracellular delivery of macromolecules comprising TRICH or fragments thereof. For example, liposome preparations containing a cell-impermeable macromolecule may promote cell fusion and intracellular delivery of the macromolecule. Alternatively, TRICH or a fragment thereof may be joined to a short cationic N-terminal portion from the HIV Tat-1 protein. Fusion proteins thus generated have been found to transduce into the cells of all tissues, including the brain, in a mouse model system (Schwarze, S.R. et al. (1999) Science 285:1569-1572).

For any compound, the therapeutically effective dose can be estimated initially either in cell culture assays, e.g., of neoplastic cells, or in animal models such as mice, rats, rabbits, dogs, monkeys, or pigs. An animal model may also be used to determine the appropriate concentration range and route of administration. Such information can then be used to determine useful doses and routes for administration in humans.

A therapeutically effective dose refers to that amount of active ingredient, for example TRICH or fragments thereof, antibodies of TRICH, and agonists, antagonists or inhibitors of TRICH, which ameliorates the symptoms or condition. Therapeutic efficacy and toxicity may be determined by standard pharmaceutical procedures in cell cultures or with experimental animals, such as by calculating the ED_{50} (the dose therapeutically effective in 50% of the population) or LD_{50} (the dose lethal to 50% of the population) statistics. The dose ratio of toxic to therapeutic effects is the therapeutic index, which can be expressed as the LD_{50}/ED_{50} ratio. Compositions which exhibit large therapeutic indices are preferred. The data obtained from cell culture assays and animal studies are used to formulate a range of dosage for human use. The dosage contained in such compositions is preferably within a range of circulating concentrations that includes the ED_{50} with little or no toxicity. The dosage varies within this range depending upon the dosage form employed, the sensitivity of the patient, and the route of administration.

The exact dosage will be determined by the practitioner, in light of factors related to the subject requiring treatment. Dosage and administration are adjusted to provide sufficient levels of the active moiety or to maintain the desired effect. Factors which may be taken into account include the severity of the disease state, the general health of the subject, the age, weight, and gender of the subject, time and frequency of administration, drug combination(s), reaction sensitivities, and response to therapy. Long-acting compositions may be administered every 3 to 4 days, every week, or biweekly depending on the half-life and clearance rate of the particular formulation.

Normal dosage amounts may vary from about 0.1 μg to 100,000 μg, up to a total dose of about 1 gram, depending upon the route of administration. Guidance as to particular dosages and methods of delivery is provided in the literature and generally available to practitioners in the art. Those skilled in the art will employ different formulations for nucleotides than for proteins or their inhibitors. Similarly, delivery of polynucleotides or polypeptides will be specific to particular cells, conditions, locations, etc.

DIAGNOSTICS

In another embodiment, antibodies which specifically bind TRICH may be used for the diagnosis of disorders characterized by expression of TRICH, or in assays to monitor patients being treated with TRICH or agonists, antagonists, or inhibitors of TRICH. Antibodies useful for diagnostic purposes may be prepared in the same manner as described above for therapeutics. Diagnostic assays for TRICH include methods which utilize the antibody and a label to detect TRICH in human body fluids or in extracts of cells or tissues. The antibodies may be used with or without modification, and may be labeled by covalent or non-covalent attachment of a reporter molecule. A wide variety of reporter molecules, several of which are described above, are known in the art and may be used.

A variety of protocols for measuring TRICH, including ELISAs, RIAs, and FACS, are known in the art and provide a basis for diagnosing altered or abnormal levels of TRICH expression. Normal or standard values for TRICH expression are established by combining body fluids or cell extracts taken from normal mammalian subjects, for example, human subjects, with antibodies to TRICH under conditions suitable for complex formation. The amount of standard complex formation may be quantitated by various methods, such as photometric means. Quantities of TRICH expressed in subject, control, and disease samples from biopsied tissues are compared with the standard values. Deviation between standard and subject values establishes the parameters for diagnosing disease.

In another embodiment of the invention, the polynucleotides encoding TRICH may be used for diagnostic purposes. The polynucleotides which may be used include oligonucleotide sequences, complementary RNA and DNA molecules, and PNAs. The polynucleotides may be used to detect and quantify gene expression in biopsied tissues in which expression of TRICH may be correlated

with disease. The diagnostic assay may be used to determine absence, presence, and excess expression of TRICH, and to monitor regulation of TRICH levels during therapeutic intervention.

In one aspect, hybridization with PCR probes which are capable of detecting polynucleotide sequences, including genomic sequences, encoding TRICH or closely related molecules may be used to identify nucleic acid sequences which encode TRICH. The specificity of the probe, whether it is made from a highly specific region, e.g., the 5' regulatory region, or from a less specific region, e.g., a conserved motif, and the stringency of the hybridization or amplification will determine whether the probe identifies only naturally occurring sequences encoding TRICH, allelic variants, or related sequences.

Probes may also be used for the detection of related sequences, and may have at least 50% sequence identity to any of the TRICH encoding sequences. The hybridization probes of the subject invention may be DNA or RNA and may be derived from the sequence of SEQ ID NO:31-60 or from genomic sequences including promoters, enhancers, and introns of the TRICH gene.

Means for producing specific hybridization probes for DNAs encoding TRICH include the cloning of polynucleotide sequences encoding TRICH or TRICH derivatives into vectors for the production of mRNA probes. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes in vitro by means of the addition of the appropriate RNA polymerases and the appropriate labeled nucleotides. Hybridization probes may be labeled by a variety of reporter groups, for example, by radionuclides such as ^{32}P or ^{35}S, or by enzymatic labels, such as alkaline phosphatase coupled to the probe via avidin/biotin coupling systems, and the like.

Polynucleotide sequences encoding TRICH may be used for the diagnosis of disorders associated with expression of TRICH. Examples of such disorders include, but are not limited to, a transport disorder such as akinesia, amyotrophic lateral sclerosis, ataxia telangiectasia, cystic fibrosis, Becker's muscular dystrophy, Bell's palsy, Charcot-Marie Tooth disease, diabetes mellitus, diabetes insipidus, diabetic neuropathy, Duchenne muscular dystrophy, hyperkalemic periodic paralysis, normokalemic periodic paralysis, Parkinson's disease, malignant hyperthermia, multidrug resistance, myasthenia gravis, myotonic dystrophy, catatonia, tardive dyskinesia, dystonias, peripheral neuropathy, cerebral neoplasms, prostate cancer, cardiac disorders associated with transport, e.g., angina, bradyarrythmia, tachyarrythmia, hypertension, Long QT syndrome, myocarditis, cardiomyopathy, nemaline myopathy, centronuclear myopathy, lipid myopathy, mitochondrial myopathy, thyrotoxic myopathy, ethanol myopathy, dermatomyositis, inclusion body myositis, infectious myositis, polymyositis, neurological disorders associated with transport, e.g., Alzheimer's disease, amnesia, bipolar disorder, dementia, depression, epilepsy, Tourette's disorder, paranoid psychoses, and schizophrenia, and other disorders associated with transport, e.g., neurofibromatosis, postherpetic neuralgia, trigeminal neuropathy, sarcoidosis, sickle cell anemia, Wilson's disease,

cataracts, infertility, pulmonary artery stenosis, sensorineural autosomal deafness, hyperglycemia, hypoglycemia, Grave's disease, goiter, Cushing's disease, Addison's disease, glucose-galactose malabsorption syndrome, hypercholesterolemia, adrenoleukodystrophy, Zellweger syndrome, Menkes disease, occipital horn syndrome, von Gierke disease, cystinuria, iminoglycinuria, Hartup disease, and Fanconi disease; a neurological disorder such as epilepsy, ischemic cerebrovascular disease, stroke, cerebral neoplasms, Alzheimer's disease, Pick's disease, Huntington's disease, dementia, Parkinson's disease and other extrapyramidal disorders, amyotrophic lateral sclerosis and other motor neuron disorders, progressive neural muscular atrophy, retinitis pigmentosa, hereditary ataxias, multiple sclerosis and other demyelinating diseases, bacterial and viral meningitis, brain abscess, subdural empyema, epidural abscess, suppurative intracranial thrombophlebitis, myelitis and radiculitis, viral central nervous system disease, prion diseases including kuru, Creutzfeldt-Jakob disease, and Gerstmann-Straussler-Scheinker syndrome, fatal familial insomnia, nutritional and metabolic diseases of the nervous system, neurofibromatosis, tuberous sclerosis, cerebelloretinal hemangioblastomatosis, encephalotrigeminal syndrome, mental retardation and other developmental disorders of the central nervous system including Down syndrome, cerebral palsy, neuroskeletal disorders, autonomic nervous system disorders, cranial nerve disorders, spinal cord diseases, muscular dystrophy and other neuromuscular disorders, peripheral nervous system disorders, dermatomyositis and polymyositis, inherited, metabolic, endocrine, and toxic myopathies, myasthenia gravis, periodic paralysis, mental disorders including mood, anxiety, and schizophrenic disorders, seasonal affective disorder (SAD), akathesia, amnesia, catatonia, diabetic neuropathy, tardive dyskinesia, dystonias, paranoid psychoses, postherpetic neuralgia, Tourette's disorder, progressive supranuclear palsy, corticobasal degeneration, and familial frontotemporal dementia; a muscle disorder such as cardiomyopathy, myocarditis, Duchenne's muscular dystrophy, Becker's muscular dystrophy, myotonic dystrophy, central core disease, nemaline myopathy, centronuclear myopathy, lipid myopathy, mitochondrial myopathy, infectious myositis, polymyositis, dermatomyositis, inclusion body myositis, thyrotoxic myopathy, ethanol myopathy, angina, anaphylactic shock, arrhythmias, asthma, cardiovascular shock, Cushing's syndrome, hypertension, hypoglycemia, myocardial infarction, migraine, pheochromocytoma, and myopathies including encephalopathy, epilepsy, Kearns-Sayre syndrome, lactic acidosis, myoclonic disorder, ophthalmoplegia, and acid maltase deficiency (AMD, also known as Pompe's disease); an immunological disorder such as acquired immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED), bronchitis, cholecystitis, contact

dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, and trauma; and a cell proliferative disorder such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and cancers including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus. The polynucleotide sequences encoding TRICH may be used in Southern or northern analysis, dot blot, or other membrane-based technologies; in PCR technologies; in dipstick, pin, and multiformat ELISA-like assays; and in microarrays utilizing fluids or tissues from patients to detect altered TRICH expression. Such qualitative or quantitative methods are well known in the art.

In a particular aspect, the nucleotide sequences encoding TRICH may be useful in assays that detect the presence of associated disorders, particularly those mentioned above. The nucleotide sequences encoding TRICH may be labeled by standard methods and added to a fluid or tissue sample from a patient under conditions suitable for the formation of hybridization complexes. After a suitable incubation period, the sample is washed and the signal is quantified and compared with a standard value. If the amount of signal in the patient sample is significantly altered in comparison to a control sample then the presence of altered levels of nucleotide sequences encoding TRICH in the sample indicates the presence of the associated disorder. Such assays may also be used to evaluate the efficacy of a particular therapeutic treatment regimen in animal studies, in clinical trials, or to monitor the treatment of an individual patient.

In order to provide a basis for the diagnosis of a disorder associated with expression of TRICH, a normal or standard profile for expression is established. This may be accomplished by combining body fluids or cell extracts taken from normal subjects, either animal or human, with a sequence, or a fragment thereof, encoding TRICH, under conditions suitable for hybridization or amplification. Standard hybridization may be quantified by comparing the values obtained from

normal subjects with values from an experiment in which a known amount of a substantially purified polynucleotide is used. Standard values obtained in this manner may be compared with values obtained from samples from patients who are symptomatic for a disorder. Deviation from standard values is used to establish the presence of a disorder.

Once the presence of a disorder is established and a treatment protocol is initiated, hybridization assays may be repeated on a regular basis to determine if the level of expression in the patient begins to approximate that which is observed in the normal subject. The results obtained from successive assays may be used to show the efficacy of treatment over a period ranging from several days to months.

With respect to cancer, the presence of an abnormal amount of transcript (either under- or overexpressed) in biopsied tissue from an individual may indicate a predisposition for the development of the disease, or may provide a means for detecting the disease prior to the appearance of actual clinical symptoms. A more definitive diagnosis of this type may allow health professionals to employ preventative measures or aggressive treatment earlier thereby preventing the development or further progression of the cancer.

Additional diagnostic uses for oligonucleotides designed from the sequences encoding TRICH may involve the use of PCR. These oligomers may be chemically synthesized, generated enzymatically, or produced in vitro. Oligomers will preferably contain a fragment of a polynucleotide encoding TRICH, or a fragment of a polynucleotide complementary to the polynucleotide encoding TRICH, and will be employed under optimized conditions for identification of a specific gene or condition. Oligomers may also be employed under less stringent conditions for detection or quantification of closely related DNA or RNA sequences.

In a particular aspect, oligonucleotide primers derived from the polynucleotide sequences encoding TRICH may be used to detect single nucleotide polymorphisms (SNPs). SNPs are substitutions, insertions and deletions that are a frequent cause of inherited or acquired genetic disease in humans. Methods of SNP detection include, but are not limited to, single-stranded conformation polymorphism (SSCP) and fluorescent SSCP (fSSCP) methods. In SSCP, oligonucleotide primers derived from the polynucleotide sequences encoding TRICH are used to amplify DNA using the polymerase chain reaction (PCR). The DNA may be derived, for example, from diseased or normal tissue, biopsy samples, bodily fluids, and the like. SNPs in the DNA cause differences in the secondary and tertiary structures of PCR products in single-stranded form, and these differences are detectable using gel electrophoresis in non-denaturing gels. In fSCCP, the oligonucleotide primers are fluorescently labeled, which allows detection of the amplimers in high-throughput equipment such as DNA sequencing machines. Additionally, sequence database analysis methods, termed in silico SNP (isSNP), are capable of identifying polymorphisms by comparing the

sequence of individual overlapping DNA fragments which assemble into a common consensus sequence. These computer-based methods filter out sequence variations due to laboratory preparation of DNA and sequencing errors using statistical models and automated analyses of DNA sequence chromatograms. In the alternative, SNPs may be detected and characterized by mass spectrometry using, for example, the high throughput MASSARRAY system (Sequenom, Inc., San Diego CA).

Methods which may also be used to quantify the expression of TRICH include radiolabeling or biotinylating nucleotides, coamplification of a control nucleic acid, and interpolating results from standard curves. (See, e.g., Melby, P.C. et al. (1993) J. Immunol. Methods 159:235-244; Duplaa, C. et al. (1993) Anal. Biochem. 212:229-236.) The speed of quantitation of multiple samples may be accelerated by running the assay in a high-throughput format where the oligomer or polynucleotide of interest is presented in various dilutions and a spectrophotometric or colorimetric response gives rapid quantitation.

In further embodiments, oligonucleotides or longer fragments derived from any of the polynucleotide sequences described herein may be used as elements on a microarray. The microarray can be used in transcript imaging techniques which monitor the relative expression levels of large numbers of genes simultaneously as described below. The microarray may also be used to identify genetic variants, mutations, and polymorphisms. This information may be used to determine gene function, to understand the genetic basis of a disorder, to diagnose a disorder, to monitor progression/regression of disease as a function of gene expression, and to develop and monitor the activities of therapeutic agents in the treatment of disease. In particular, this information may be used to develop a pharmacogenomic profile of a patient in order to select the most appropriate and effective treatment regimen for that patient. For example, therapeutic agents which are highly effective and display the fewest side effects may be selected for a patient based on his/her pharmacogenomic profile.

In another embodiment, TRICH, fragments of TRICH, or antibodies specific for TRICH may be used as elements on a microarray. The microarray may be used to monitor or measure protein-protein interactions, drug-target interactions, and gene expression profiles, as described above.

A particular embodiment relates to the use of the polynucleotides of the present invention to generate a transcript image of a tissue or cell type. A transcript image represents the global pattern of gene expression by a particular tissue or cell type. Global gene expression patterns are analyzed by quantifying the number of expressed genes and their relative abundance under given conditions and at a given time. (See Seilhamer et al., "Comparative Gene Transcript Analysis," U.S. Patent Number 5,840,484, expressly incorporated by reference herein.) Thus a transcript image may be generated by hybridizing the polynucleotides of the present invention or their complements to the totality of transcripts or reverse transcripts of a particular tissue or cell type. In one embodiment, the

hybridization takes place in high-throughput format, wherein the polynucleotides of the present invention or their complements comprise a subset of a plurality of elements on a microarray. The resultant transcript image would provide a profile of gene activity.

Transcript images may be generated using transcripts isolated from tissues, cell lines, biopsies, or other biological samples. The transcript image may thus reflect gene expression in vivo, as in the case of a tissue or biopsy sample, or in vitro, as in the case of a cell line.

Transcript images which profile the expression of the polynucleotides of the present invention may also be used in conjunction with in vitro model systems and preclinical evaluation of pharmaceuticals, as well as toxicological testing of industrial and naturally-occurring environmental compounds. All compounds induce characteristic gene expression patterns, frequently termed molecular fingerprints or toxicant signatures, which are indicative of mechanisms of action and toxicity (Nuwaysir, E.F. et al. (1999) Mol. Carcinog. 24:153-159; Steiner, S. and N.L. Anderson (2000) Toxicol. Lett. 112-113:467-471, expressly incorporated by reference herein). If a test compound has a signature similar to that of a compound with known toxicity, it is likely to share those toxic properties. These fingerprints or signatures are most useful and refined when they contain expression information from a large number of genes and gene families. Ideally, a genome-wide measurement of expression provides the highest quality signature. Even genes whose expression is not altered by any tested compounds are important as well, as the levels of expression of these genes are used to normalize the rest of the expression data. The normalization procedure is useful for comparison of expression data after treatment with different compounds. While the assignment of gene function to elements of a toxicant signature aids in interpretation of toxicity mechanisms, knowledge of gene function is not necessary for the statistical matching of signatures which leads to prediction of toxicity. (See, for example, Press Release 00-02 from the National Institute of Environmental Health Sciences, released February 29, 2000, available at http://www.niehs.nih.gov/oc/news/toxchip.htm.) Therefore, it is important and desirable in toxicological screening using toxicant signatures to include all expressed gene sequences.

In one embodiment, the toxicity of a test compound is assessed by treating a biological sample containing nucleic acids with the test compound. Nucleic acids that are expressed in the treated biological sample are hybridized with one or more probes specific to the polynucleotides of the present invention, so that transcript levels corresponding to the polynucleotides of the present invention may be quantified. The transcript levels in the treated biological sample are compared with levels in an untreated biological sample. Differences in the transcript levels between the two samples are indicative of a toxic response caused by the test compound in the treated sample.

Another particular embodiment relates to the use of the polypeptide sequences of the present invention to analyze the proteome of a tissue or cell type. The term proteome refers to the global

pattern of protein expression in a particular tissue or cell type. Each protein component of a proteome can be subjected individually to further analysis. Proteome expression patterns, or profiles, are analyzed by quantifying the number of expressed proteins and their relative abundance under given conditions and at a given time. A profile of a cell's proteome may thus be generated by separating and analyzing the polypeptides of a particular tissue or cell type. In one embodiment, the separation is achieved using two-dimensional gel electrophoresis, in which proteins from a sample are separated by isoelectric focusing in the first dimension, and then according to molecular weight by sodium dodecyl sulfate slab gel electrophoresis in the second dimension (Steiner and Anderson, supra). The proteins are visualized in the gel as discrete and uniquely positioned spots, typically by staining the gel with an agent such as Coomassie Blue or silver or fluorescent stains. The optical density of each protein spot is generally proportional to the level of the protein in the sample. The optical densities of equivalently positioned protein spots from different samples, for example, from biological samples either treated or untreated with a test compound or therapeutic agent, are compared to identify any changes in protein spot density related to the treatment. The proteins in the spots are partially sequenced using, for example, standard methods employing chemical or enzymatic cleavage followed by mass spectrometry. The identity of the protein in a spot may be determined by comparing its partial sequence, preferably of at least 5 contiguous amino acid residues, to the polypeptide sequences of the present invention. In some cases, further sequence data may be obtained for definitive protein identification.

A proteomic profile may also be generated using antibodies specific for TRICH to quantify the levels of TRICH expression. In one embodiment, the antibodies are used as elements on a microarray, and protein expression levels are quantified by exposing the microarray to the sample and detecting the levels of protein bound to each array element (Lueking, A. et al. (1999) Anal. Biochem. 270:103-111; Mendoze, L.G. et al. (1999) Biotechniques 27:778-788). Detection may be performed by a variety of methods known in the art, for example, by reacting the proteins in the sample with a thiol- or amino-reactive fluorescent compound and detecting the amount of fluorescence bound at each array element.

Toxicant signatures at the proteome level are also useful for toxicological screening, and should be analyzed in parallel with toxicant signatures at the transcript level. There is a poor correlation between transcript and protein abundances for some proteins in some tissues (Anderson, N.L. and J. Seilhamer (1997) Electrophoresis 18:533-537), so proteome toxicant signatures may be useful in the analysis of compounds which do not significantly affect the transcript image, but which alter the proteomic profile. In addition, the analysis of transcripts in body fluids is difficult, due to rapid degradation of mRNA, so proteomic profiling may be more reliable and informative in such cases.

In another embodiment, the toxicity of a test compound is assessed by treating a biological sample containing proteins with the test compound. Proteins that are expressed in the treated biological sample are separated so that the amount of each protein can be quantified. The amount of each protein is compared to the amount of the corresponding protein in an untreated biological sample. A difference in the amount of protein between the two samples is indicative of a toxic response to the test compound in the treated sample. Individual proteins are identified by sequencing the amino acid residues of the individual proteins and comparing these partial sequences to the polypeptides of the present invention.

In another embodiment, the toxicity of a test compound is assessed by treating a biological sample containing proteins with the test compound. Proteins from the biological sample are incubated with antibodies specific to the polypeptides of the present invention. The amount of protein recognized by the antibodies is quantified. The amount of protein in the treated biological sample is compared with the amount in an untreated biological sample. A difference in the amount of protein between the two samples is indicative of a toxic response to the test compound in the treated sample.

Microarrays may be prepared, used, and analyzed using methods known in the art. (See, e.g., Brennan, T.M. et al. (1995) U.S. Patent No. 5,474,796; Schena, M. et al. (1996) Proc. Natl. Acad. Sci. USA 93:10614-10619; Baldeschweiler et al. (1995) PCT application WO95/251116; Shalon, D. et al. (1995) PCT application WO95/35505; Heller, R.A. et al. (1997) Proc. Natl. Acad. Sci. USA 94:2150-2155; and Heller, M.J. et al. (1997) U.S. Patent No. 5,605,662.) Various types of microarrays are well known and thoroughly described in DNA Microarrays: A Practical Approach, M. Schena, ed. (1999) Oxford University Press, London, hereby expressly incorporated by reference.

In another embodiment of the invention, nucleic acid sequences encoding TRICH may be used to generate hybridization probes useful in mapping the naturally occurring genomic sequence. Either coding or noncoding sequences may be used, and in some instances, noncoding sequences may be preferable over coding sequences. For example, conservation of a coding sequence among members of a multi-gene family may potentially cause undesired cross hybridization during chromosomal mapping. The sequences may be mapped to a particular chromosome, to a specific region of a chromosome, or to artificial chromosome constructions, e.g., human artificial chromosomes (HACs), yeast artificial chromosomes (YACs), bacterial artificial chromosomes (BACs), bacterial P1 constructions, or single chromosome cDNA libraries. (See, e.g., Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355; Price, C.M. (1993) Blood Rev. 7:127-134; and Trask, B.J. (1991) Trends Genet. 7:149-154.) Once mapped, the nucleic acid sequences of the invention may be used to develop genetic linkage maps, for example, which correlate the inheritance of a disease state with the inheritance of a particular chromosome region or restriction fragment length polymorphism

(RFLP). (See, for example, Lander, E.S. and D. Botstein (1986) Proc. Natl. Acad. Sci. USA 83:7353-7357.)

Fluorescent in situ hybridization (FISH) may be correlated with other physical and genetic map data. (See, e.g., Heinz-Ulrich, et al. (1995) in Meyers, supra, pp. 965-968.) Examples of genetic map data can be found in various scientific journals or at the Online Mendelian Inheritance in Man (OMIM) World Wide Web site. Correlation between the location of the gene encoding TRICH on a physical map and a specific disorder, or a predisposition to a specific disorder, may help define the region of DNA associated with that disorder and thus may further positional cloning efforts.

In situ hybridization of chromosomal preparations and physical mapping techniques, such as linkage analysis using established chromosomal markers, may be used for extending genetic maps. Often the placement of a gene on the chromosome of another mammalian species, such as mouse, may reveal associated markers even if the exact chromosomal locus is not known. This information is valuable to investigators searching for disease genes using positional cloning or other gene discovery techniques. Once the gene or genes responsible for a disease or syndrome have been crudely localized by genetic linkage to a particular genomic region, e.g., ataxia-telangiectasia to 11q22-23, any sequences mapping to that area may represent associated or regulatory genes for further investigation. (See, e.g., Gatti, R.A. et al. (1988) Nature 336:577-580.) The nucleotide sequence of the instant invention may also be used to detect differences in the chromosomal location due to translocation, inversion, etc., among normal, carrier, or affected individuals.

In another embodiment of the invention, TRICH, its catalytic or immunogenic fragments, or oligopeptides thereof can be used for screening libraries of compounds in any of a variety of drug screening techniques. The fragment employed in such screening may be free in solution, affixed to a solid support, borne on a cell surface, or located intracellularly. The formation of binding complexes between TRICH and the agent being tested may be measured.

Another technique for drug screening provides for high throughput screening of compounds having suitable binding affinity to the protein of interest. (See, e.g., Geysen, et al. (1984) PCT application WO84/03564.) In this method, large numbers of different small test compounds are synthesized on a solid substrate. The test compounds are reacted with TRICH, or fragments thereof, and washed. Bound TRICH is then detected by methods well known in the art. Purified TRICH can also be coated directly onto plates for use in the aforementioned drug screening techniques. Alternatively, non-neutralizing antibodies can be used to capture the peptide and immobilize it on a solid support.

In another embodiment, one may use competitive drug screening assays in which neutralizing antibodies capable of binding TRICH specifically compete with a test compound for binding TRICH. In this manner, antibodies can be used to detect the presence of any peptide which shares one or more

antigenic determinants with TRICH.

In additional embodiments, the nucleotide sequences which encode TRICH may be used in any molecular biology techniques that have yet to be developed, provided the new techniques rely on properties of nucleotide sequences that are currently known, including, but not limited to, such properties as the triplet genetic code and specific base pair interactions.

Without further elaboration, it is believed that one skilled in the art can, using the preceding description, utilize the present invention to its fullest extent. The following embodiments are, therefore, to be construed as merely illustrative, and not limitative of the remainder of the disclosure in any way whatsoever.

The disclosures of all patents, applications, and publications mentioned above and below, including U.S. Ser. No. 60/223,269, U.S. Ser. No. 60/224,456, U.S. Ser. No. 60/226,410, U.S. Ser. No. 60/228,140, U.S. Ser. No. 60/230,067, and U.S. Ser. No. 60/231,434, are hereby expressly incorporated by reference.

EXAMPLES

I. Construction of cDNA Libraries

Incyte cDNAs were derived from cDNA libraries described in the LIFESEQ GOLD database (Incyte Genomics, Palo Alto CA) and shown in Table 4, column 5. Some tissues were homogenized and lysed in guanidinium isothiocyanate, while others were homogenized and lysed in phenol or in a suitable mixture of denaturants, such as TRIZOL (Life Technologies), a monophasic solution of phenol and guanidine isothiocyanate. The resulting lysates were centrifuged over CsCl cushions or extracted with chloroform. RNA was precipitated from the lysates with either isopropanol or sodium acetate and ethanol, or by other routine methods.

Phenol extraction and precipitation of RNA were repeated as necessary to increase RNA purity. In some cases, RNA was treated with DNase. For most libraries, poly(A)+ RNA was isolated using oligo d(T)-coupled paramagnetic particles (Promega), OLIGOTEX latex particles (QIAGEN, Chatsworth CA), or an OLIGOTEX mRNA purification kit (QIAGEN). Alternatively, RNA was isolated directly from tissue lysates using other RNA isolation kits, e.g., the POLY(A)PURE mRNA purification kit (Ambion, Austin TX).

In some cases, Stratagene was provided with RNA and constructed the corresponding cDNA libraries. Otherwise, cDNA was synthesized and cDNA libraries were constructed with the UNIZAP vector system (Stratagene) or SUPERSCRIPT plasmid system (Life Technologies), using the recommended procedures or similar methods known in the art. (See, e.g., Ausubel, 1997, supra, units 5.1-6.6.) Reverse transcription was initiated using oligo d(T) or random primers. Synthetic oligonucleotide adapters were ligated to double stranded cDNA, and the cDNA was digested with the

appropriate restriction enzyme or enzymes. For most libraries, the cDNA was size-selected (300-1000 bp) using SEPHACRYL S1000, SEPHAROSE CL2B, or SEPHAROSE CL4B column chromatography (Amersham Pharmacia Biotech) or preparative agarose gel electrophoresis. cDNAs were ligated into compatible restriction enzyme sites of the polylinker of a suitable plasmid, e.g., PBLUESCRIPT plasmid (Stratagene), PSPORT1 plasmid (Life Technologies), PCDNA2.1 plasmid (Invitrogen, Carlsbad CA), PBK-CMV plasmid (Stratagene), or pINCY (Incyte Genomics, Palo Alto CA), or derivatives thereof. Recombinant plasmids were transformed into competent E. coli cells including XL1-Blue, XL1-BlueMRF, or SOLR from Stratagene or DH5α, DH10B, or ElectroMAX DH10B from Life Technologies.

II. Isolation of cDNA Clones

Plasmids obtained as described in Example I were recovered from host cells by in vivo excision using the UNIZAP vector system (Stratagene) or by cell lysis. Plasmids were purified using at least one of the following: a Magic or WIZARD Minipreps DNA purification system (Promega); an AGTC Miniprep purification kit (Edge Biosystems, Gaithersburg MD); and QIAWELL 8 Plasmid, QIAWELL 8 Plus Plasmid, QIAWELL 8 Ultra Plasmid purification systems or the R.E.A.L. PREP 96 plasmid purification kit from QIAGEN. Following precipitation, plasmids were resuspended in 0.1 ml of distilled water and stored, with or without lyophilization, at 4°C.

Alternatively, plasmid DNA was amplified from host cell lysates using direct link PCR in a high-throughput format (Rao, V.B. (1994) Anal. Biochem. 216:1-14). Host cell lysis and thermal cycling steps were carried out in a single reaction mixture. Samples were processed and stored in 384-well plates, and the concentration of amplified plasmid DNA was quantified fluorometrically using PICOGREEN dye (Molecular Probes, Eugene OR) and a FLUOROSKAN II fluorescence scanner (Labsystems Oy, Helsinki, Finland).

III. Sequencing and Analysis

Incyte cDNA recovered in plasmids as described in Example II were sequenced as follows. Sequencing reactions were processed using standard methods or high-throughput instrumentation such as the ABI CATALYST 800 (Applied Biosystems) thermal cycler or the PTC-200 thermal cycler (MJ Research) in conjunction with the HYDRA microdispenser (Robbins Scientific) or the MICROLAB 2200 (Hamilton) liquid transfer system. cDNA sequencing reactions were prepared using reagents provided by Amersham Pharmacia Biotech or supplied in ABI sequencing kits such as the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Applied Biosystems). Electrophoretic separation of cDNA sequencing reactions and detection of labeled polynucleotides were carried out using the MEGABACE 1000 DNA sequencing system (Molecular Dynamics); the ABI PRISM 373 or 377 sequencing system (Applied Biosystems) in conjunction with standard ABI protocols and base calling software; or other sequence analysis systems known in the art. Reading

frames within the cDNA sequences were identified using standard methods (reviewed in Ausubel, 1997, supra, unit 7.7). Some of the cDNA sequences were selected for extension using the techniques disclosed in Example VIII.

The polynucleotide sequences derived from Incyte cDNAs were validated by removing vector, linker, and poly(A) sequences and by masking ambiguous bases, using algorithms and programs based on BLAST, dynamic programming, and dinucleotide nearest neighbor analysis. The Incyte cDNA sequences or translations thereof were then queried against a selection of public databases such as the GenBank primate, rodent, mammalian, vertebrate, and eukaryote databases, and BLOCKS, PRINTS, DOMO, PRODOM, and hidden Markov model (HMM)-based protein family databases such as PFAM. (HMM is a probabilistic approach which analyzes consensus primary structures of gene families. See, for example, Eddy, S.R. (1996) Curr. Opin. Struct. Biol. 6:361-365.) The queries were performed using programs based on BLAST, FASTA, BLIMPS, and HMMER. The Incyte cDNA sequences were assembled to produce full length polynucleotide sequences. Alternatively, GenBank cDNAs, GenBank ESTs, stitched sequences, stretched sequences, or Genscan-predicted coding sequences (see Examples IV and V) were used to extend Incyte cDNA assemblages to full length. Assembly was performed using programs based on Phred, Phrap, and Consed, and cDNA assemblages were screened for open reading frames using programs based on GeneMark, BLAST, and FASTA. The full length polynucleotide sequences were translated to derive the corresponding full length polypeptide sequences. Alternatively, a polypeptide of the invention may begin at any of the methionine residues of the full length translated polypeptide. Full length polypeptide sequences were subsequently analyzed by querying against databases such as the GenBank protein databases (genpept), SwissProt, BLOCKS, PRINTS, DOMO, PRODOM, Prosite, and hidden Markov model (HMM)-based protein family databases such as PFAM. Full length polynucleotide sequences are also analyzed using MACDNASIS PRO software (Hitachi Software Engineering, South San Francisco CA) and LASERGENE software (DNASTAR). Polynucleotide and polypeptide sequence alignments are generated using default parameters specified by the CLUSTAL algorithm as incorporated into the MEGALIGN multisequence alignment program (DNASTAR), which also calculates the percent identity between aligned sequences.

Table 7 summarizes the tools, programs, and algorithms used for the analysis and assembly of Incyte cDNA and full length sequences and provides applicable descriptions, references, and threshold parameters. The first column of Table 7 shows the tools, programs, and algorithms used, the second column provides brief descriptions thereof, the third column presents appropriate references, all of which are incorporated by reference herein in their entirety, and the fourth column presents, where applicable, the scores, probability values, and other parameters used to evaluate the strength of a match between two sequences (the higher the score or the lower the probability value,

the greater the identity between two sequences).

The programs described above for the assembly and analysis of full length polynucleotide and polypeptide sequences were also used to identify polynucleotide sequence fragments from SEQ ID NO:31-60. Fragments from about 20 to about 4000 nucleotides which are useful in hybridization and amplification technologies are described in Table 4, column 4.

IV. Identification and Editing of Coding Sequences from Genomic DNA

Putative transporters and ion channels were initially identified by running the Genscan gene identification program against public genomic sequence databases (e.g., gbpri and gbhtg). Genscan is a general-purpose gene identification program which analyzes genomic DNA sequences from a variety of organisms (See Burge, C. and S. Karlin (1997) J. Mol. Biol. 268:78-94, and Burge, C. and S. Karlin (1998) Curr. Opin. Struct. Biol. 8:346-354). The program concatenates predicted exons to form an assembled cDNA sequence extending from a methionine to a stop codon. The output of Genscan is a FASTA database of polynucleotide and polypeptide sequences. The maximum range of sequence for Genscan to analyze at once was set to 30 kb. To determine which of these Genscan predicted cDNA sequences encode transporters and ion channels, the encoded polypeptides were analyzed by querying against PFAM models for transporters and ion channels. Potential transporters and ion channels were also identified by homology to Incyte cDNA sequences that had been annotated as transporters and ion channels. These selected Genscan-predicted sequences were then compared by BLAST analysis to the genpept and gbpri public databases. Where necessary, the Genscan-predicted sequences were then edited by comparison to the top BLAST hit from genpept to correct errors in the sequence predicted by Genscan, such as extra or omitted exons. BLAST analysis was also used to find any Incyte cDNA or public cDNA coverage of the Genscan-predicted sequences, thus providing evidence for transcription. When Incyte cDNA coverage was available, this information was used to correct or confirm the Genscan predicted sequence. Full length polynucleotide sequences were obtained by assembling Genscan-predicted coding sequences with Incyte cDNA sequences and/or public cDNA sequences using the assembly process described in Example III. Alternatively, full length polynucleotide sequences were derived entirely from edited or unedited Genscan-predicted coding sequences.

V. Assembly of Genomic Sequence Data with cDNA Sequence Data

"Stitched" Sequences

Partial cDNA sequences were extended with exons predicted by the Genscan gene identification program described in Example IV. Partial cDNAs assembled as described in Example III were mapped to genomic DNA and parsed into clusters containing related cDNAs and Genscan exon predictions from one or more genomic sequences. Each cluster was analyzed using an algorithm based on graph theory and dynamic programming to integrate cDNA and genomic information,

generating possible splice variants that were subsequently confirmed, edited, or extended to create a full length sequence. Sequence intervals in which the entire length of the interval was present on more than one sequence in the cluster were identified, and intervals thus identified were considered to be equivalent by transitivity. For example, if an interval was present on a cDNA and two genomic sequences, then all three intervals were considered to be equivalent. This process allows unrelated but consecutive genomic sequences to be brought together, bridged by cDNA sequence. Intervals thus identified were then "stitched" together by the stitching algorithm in the order that they appear along their parent sequences to generate the longest possible sequence, as well as sequence variants. Linkages between intervals which proceed along one type of parent sequence (cDNA to cDNA or genomic sequence to genomic sequence) were given preference over linkages which change parent type (cDNA to genomic sequence). The resultant stitched sequences were translated and compared by BLAST analysis to the genpept and gbpri public databases. Incorrect exons predicted by Genscan were corrected by comparison to the top BLAST hit from genpept. Sequences were further extended with additional cDNA sequences, or by inspection of genomic DNA, when necessary.

"Stretched" Sequences

Partial DNA sequences were extended to full length with an algorithm based on BLAST analysis. First, partial cDNAs assembled as described in Example III were queried against public databases such as the GenBank primate, rodent, mammalian, vertebrate, and eukaryote databases using the BLAST program. The nearest GenBank protein homolog was then compared by BLAST analysis to either Incyte cDNA sequences or GenScan exon predicted sequences described in Example IV. A chimeric protein was generated by using the resultant high-scoring segment pairs (HSPs) to map the translated sequences onto the GenBank protein homolog. Insertions or deletions may occur in the chimeric protein with respect to the original GenBank protein homolog. The GenBank protein homolog, the chimeric protein, or both were used as probes to search for homologous genomic sequences from the public human genome databases. Partial DNA sequences were therefore "stretched" or extended by the addition of homologous genomic sequences. The resultant stretched sequences were examined to determine whether it contained a complete gene.

VI. Chromosomal Mapping of TRICH Encoding Polynucleotides

The sequences which were used to assemble SEQ ID NO:31-60 were compared with sequences from the Incyte LIFESEQ database and public domain databases using BLAST and other implementations of the Smith-Waterman algorithm. Sequences from these databases that matched SEQ ID NO:31-60 were assembled into clusters of contiguous and overlapping sequences using assembly algorithms such as Phrap (Table 7). Radiation hybrid and genetic mapping data available from public resources such as the Stanford Human Genome Center (SHGC), Whitehead Institute for Genome Research (WIGR), and Généthon were used to determine if any of the clustered sequences

had been previously mapped. Inclusion of a mapped sequence in a cluster resulted in the assignment of all sequences of that cluster, including its particular SEQ ID NO:, to that map location.

Map locations are represented by ranges, or intervals, of human chromosomes. The map position of an interval, in centiMorgans, is measured relative to the terminus of the chromosome's p-arm. (The centiMorgan (cM) is a unit of measurement based on recombination frequencies between chromosomal markers. On average, 1 cM is roughly equivalent to 1 megabase (Mb) of DNA in humans, although this can vary widely due to hot and cold spots of recombination.) The cM distances are based on genetic markers mapped by Généthon which provide boundaries for radiation hybrid markers whose sequences were included in each of the clusters. Human genome maps and other resources available to the public, such as the NCBI "GeneMap'99" World Wide Web site (http://www.ncbi.nlm.nih.gov/genemap/), can be employed to determine if previously identified disease genes map within or in proximity to the intervals indicated above.

VII. Analysis of Polynucleotide Expression

Northern analysis is a laboratory technique used to detect the presence of a transcript of a gene and involves the hybridization of a labeled nucleotide sequence to a membrane on which RNAs from a particular cell type or tissue have been bound. (See, e.g., Sambrook, supra, ch. 7; Ausubel (1995) supra, ch. 4 and 16.)

Analogous computer techniques applying BLAST were used to search for identical or related molecules in cDNA databases such as GenBank or LIFESEQ (Incyte Genomics). This analysis is much faster than multiple membrane-based hybridizations. In addition, the sensitivity of the computer search can be modified to determine whether any particular match is categorized as exact or similar. The basis of the search is the product score, which is defined as:

$$\frac{\text{BLAST Score x Percent Identity}}{\text{5 x minimum \{length(Seq. 1), length(Seq. 2)\}}}$$

The product score takes into account both the degree of similarity between two sequences and the length of the sequence match. The product score is a normalized value between 0 and 100, and is calculated as follows: the BLAST score is multiplied by the percent nucleotide identity and the product is divided by (5 times the length of the shorter of the two sequences). The BLAST score is calculated by assigning a score of +5 for every base that matches in a high-scoring segment pair (HSP), and -4 for every mismatch. Two sequences may share more than one HSP (separated by gaps). If there is more than one HSP, then the pair with the highest BLAST score is used to calculate the product score. The product score represents a balance between fractional overlap and quality in a BLAST alignment. For example, a product score of 100 is produced only for 100% identity over the

entire length of the shorter of the two sequences being compared. A product score of 70 is produced either by 100% identity and 70% overlap at one end, or by 88% identity and 100% overlap at the other. A product score of 50 is produced either by 100% identity and 50% overlap at one end, or 79% identity and 100% overlap.

Alternatively, polynucleotide sequences encoding TRICH are analyzed with respect to the tissue sources from which they were derived. For example, some full length sequences are assembled, at least in part, with overlapping Incyte cDNA sequences (see Example III). Each cDNA sequence is derived from a cDNA library constructed from a human tissue. Each human tissue is classified into one of the following organ/tissue categories: cardiovascular system; connective tissue; digestive system; embryonic structures; endocrine system; exocrine glands; genitalia, female; genitalia, male; germ cells; hemic and immune system; liver; musculoskeletal system; nervous system; pancreas; respiratory system; sense organs; skin; stomatognathic system; unclassified/mixed; or urinary tract. The number of libraries in each category is counted and divided by the total number of libraries across all categories. Similarly, each human tissue is classified into one of the following disease/condition categories: cancer, cell line, developmental, inflammation, neurological, trauma, cardiovascular, pooled, and other, and the number of libraries in each category is counted and divided by the total number of libraries across all categories. The resulting percentages reflect the tissue- and disease-specific expression of cDNA encoding TRICH. cDNA sequences and cDNA library/tissue information are found in the LIFESEQ GOLD database (Incyte Genomics, Palo Alto CA).

VIII. Extension of TRICH Encoding Polynucleotides

Full length polynucleotide sequences were also produced by extension of an appropriate fragment of the full length molecule using oligonucleotide primers designed from this fragment. One primer was synthesized to initiate 5' extension of the known fragment, and the other primer was synthesized to initiate 3' extension of the known fragment. The initial primers were designed using OLIGO 4.06 software (National Biosciences), or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the target sequence at temperatures of about 68°C to about 72°C. Any stretch of nucleotides which would result in hairpin structures and primer-primer dimerizations was avoided.

Selected human cDNA libraries were used to extend the sequence. If more than one extension was necessary or desired, additional or nested sets of primers were designed.

High fidelity amplification was obtained by PCR using methods well known in the art. PCR was performed in 96-well plates using the PTC-200 thermal cycler (MJ Research, Inc.). The reaction mix contained DNA template, 200 nmol of each primer, reaction buffer containing Mg^{2+}, $(NH_4)_2SO_4$, and 2-mercaptoethanol, Taq DNA polymerase (Amersham Pharmacia Biotech), ELONGASE enzyme (Life Technologies), and Pfu DNA polymerase (Stratagene), with the following parameters for primer

pair PCI A and PCI B: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C. In the alternative, the parameters for primer pair T7 and SK+ were as follows: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 57°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C.

The concentration of DNA in each well was determined by dispensing 100 μl PICOGREEN quantitation reagent (0.25% (v/v) PICOGREEN; Molecular Probes, Eugene OR) dissolved in 1X TE and 0.5 μl of undiluted PCR product into each well of an opaque fluorimeter plate (Corning Costar, Acton MA), allowing the DNA to bind to the reagent. The plate was scanned in a Fluoroskan II (Labsystems Oy, Helsinki, Finland) to measure the fluorescence of the sample and to quantify the concentration of DNA. A 5 μl to 10 μl aliquot of the reaction mixture was analyzed by electrophoresis on a 1 % agarose gel to determine which reactions were successful in extending the sequence.

The extended nucleotides were desalted and concentrated, transferred to 384-well plates, digested with CviJI cholera virus endonuclease (Molecular Biology Research, Madison WI), and sonicated or sheared prior to religation into pUC 18 vector (Amersham Pharmacia Biotech). For shotgun sequencing, the digested nucleotides were separated on low concentration (0.6 to 0.8%) agarose gels, fragments were excised, and agar digested with Agar ACE (Promega). Extended clones were religated using T4 ligase (New England Biolabs, Beverly MA) into pUC 18 vector (Amersham Pharmacia Biotech), treated with Pfu DNA polymerase (Stratagene) to fill-in restriction site overhangs, and transfected into competent E. coli cells. Transformed cells were selected on antibiotic-containing media, and individual colonies were picked and cultured overnight at 37°C in 384-well plates in LB/2x carb liquid media.

The cells were lysed, and DNA was amplified by PCR using Taq DNA polymerase (Amersham Pharmacia Biotech) and Pfu DNA polymerase (Stratagene) with the following parameters: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 72°C, 2 min; Step 5: steps 2, 3, and 4 repeated 29 times; Step 6: 72°C, 5 min; Step 7: storage at 4°C. DNA was quantified by PICOGREEN reagent (Molecular Probes) as described above. Samples with low DNA recoveries were reamplified using the same conditions as described above. Samples were diluted with 20% dimethysulfoxide (1:2, v/v), and sequenced using DYENAMIC energy transfer sequencing primers and the DYENAMIC DIRECT kit (Amersham Pharmacia Biotech) or the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Applied Biosystems).

In like manner, full length polynucleotide sequences are verified using the above procedure or are used to obtain 5' regulatory sequences using the above procedure along with oligonucleotides designed for such extension, and an appropriate genomic library.

IX. Labeling and Use of Individual Hybridization Probes

Hybridization probes derived from SEQ ID NO:31-60 are employed to screen cDNAs, genomic DNAs, or mRNAs. Although the labeling of oligonucleotides, consisting of about 20 base pairs, is specifically described, essentially the same procedure is used with larger nucleotide fragments. Oligonucleotides are designed using state-of-the-art software such as OLIGO 4.06 software (National Biosciences) and labeled by combining 50 pmol of each oligomer, 250 μCi of [γ-^{32}P] adenosine triphosphate (Amersham Pharmacia Biotech), and T4 polynucleotide kinase (DuPont NEN, Boston MA). The labeled oligonucleotides are substantially purified using a SEPHADEX G-25 superfine size exclusion dextran bead column (Amersham Pharmacia Biotech). An aliquot containing 10^7 counts per minute of the labeled probe is used in a typical membrane-based hybridization analysis of human genomic DNA digested with one of the following endonucleases: Ase I, Bgl II, Eco RI, Pst I, Xba I, or Pvu II (DuPont NEN).

The DNA from each digest is fractionated on a 0.7% agarose gel and transferred to nylon membranes (Nytran Plus, Schleicher & Schuell, Durham NH). Hybridization is carried out for 16 hours at 40°C. To remove nonspecific signals, blots are sequentially washed at room temperature under conditions of up to, for example, 0.1 x saline sodium citrate and 0.5% sodium dodecyl sulfate. Hybridization patterns are visualized using autoradiography or an alternative imaging means and compared.

X. Microarrays

The linkage or synthesis of array elements upon a microarray can be achieved utilizing photolithography, piezoelectric printing (ink-jet printing, See, e.g., Baldeschweiler, supra.), mechanical microspotting technologies, and derivatives thereof. The substrate in each of the aforementioned technologies should be uniform and solid with a non-porous surface (Schena (1999), supra). Suggested substrates include silicon, silica, glass slides, glass chips, and silicon wafers. Alternatively, a procedure analogous to a dot or slot blot may also be used to arrange and link elements to the surface of a substrate using thermal, UV, chemical, or mechanical bonding procedures. A typical array may be produced using available methods and machines well known to those of ordinary skill in the art and may contain any appropriate number of elements. (See, e.g., Schena, M. et al. (1995) Science 270:467-470; Shalon, D. et al. (1996) Genome Res. 6:639-645; Marshall, A. and J. Hodgson (1998) Nat. Biotechnol. 16:27-31.)

Full length cDNAs, Expressed Sequence Tags (ESTs), or fragments or oligomers thereof may comprise the elements of the microarray. Fragments or oligomers suitable for hybridization can be selected using software well known in the art such as LASERGENE software (DNASTAR). The array elements are hybridized with polynucleotides in a biological sample. The polynucleotides in the biological sample are conjugated to a fluorescent label or other molecular tag for ease of detection.

After hybridization, nonhybridized nucleotides from the biological sample are removed, and a fluorescence scanner is used to detect hybridization at each array element. Alternatively, laser desorbtion and mass spectrometry may be used for detection of hybridization. The degree of complementarity and the relative abundance of each polynucleotide which hybridizes to an element on the microarray may be assessed. In one embodiment, microarray preparation and usage is described in detail below.

<u>Tissue or Cell Sample Preparation</u>

Total RNA is isolated from tissue samples using the guanidinium thiocyanate method and poly(A)$^+$ RNA is purified using the oligo-(dT) cellulose method. Each poly(A)$^+$ RNA sample is reverse transcribed using MMLV reverse-transcriptase, 0.05 pg/μl oligo-(dT) primer (21mer), 1X first strand buffer, 0.03 units/μl RNase inhibitor, 500 μM dATP, 500 μM dGTP, 500 μM dTTP, 40 μM dCTP, 40 μM dCTP-Cy3 (BDS) or dCTP-Cy5 (Amersham Pharmacia Biotech). The reverse transcription reaction is performed in a 25 ml volume containing 200 ng poly(A)$^+$ RNA with GEMBRIGHT kits (Incyte). Specific control poly(A)$^+$ RNAs are synthesized by <u>in vitro</u> transcription from non-coding yeast genomic DNA. After incubation at 37°C for 2 hr, each reaction sample (one with Cy3 and another with Cy5 labeling) is treated with 2.5 ml of 0.5M sodium hydroxide and incubated for 20 minutes at 85°C to the stop the reaction and degrade the RNA. Samples are purified using two successive CHROMA SPIN 30 gel filtration spin columns (CLONTECH Laboratories, Inc. (CLONTECH), Palo Alto CA) and after combining, both reaction samples are ethanol precipitated using 1 ml of glycogen (1 mg/ml), 60 ml sodium acetate, and 300 ml of 100% ethanol. The sample is then dried to completion using a SpeedVAC (Savant Instruments Inc., Holbrook NY) and resuspended in 14 μl 5X SSC/0.2% SDS.

<u>Microarray Preparation</u>

Sequences of the present invention are used to generate array elements. Each array element is amplified from bacterial cells containing vectors with cloned cDNA inserts. PCR amplification uses primers complementary to the vector sequences flanking the cDNA insert. Array elements are amplified in thirty cycles of PCR from an initial quantity of 1-2 ng to a final quantity greater than 5 μg. Amplified array elements are then purified using SEPHACRYL-400 (Amersham Pharmacia Biotech).

Purified array elements are immobilized on polymer-coated glass slides. Glass microscope slides (Corning) are cleaned by ultrasound in 0.1% SDS and acetone, with extensive distilled water washes between and after treatments. Glass slides are etched in 4% hydrofluoric acid (VWR Scientific Products Corporation (VWR), West Chester PA), washed extensively in distilled water, and coated with 0.05% aminopropyl silane (Sigma) in 95% ethanol. Coated slides are cured in a

110°C oven.

Array elements are applied to the coated glass substrate using a procedure described in US Patent No. 5,807,522, incorporated herein by reference. 1 μl of the array element DNA, at an average concentration of 100 ng/μl, is loaded into the open capillary printing element by a high-speed robotic apparatus. The apparatus then deposits about 5 nl of array element sample per slide.

Microarrays are UV-crosslinked using a STRATALINKER UV-crosslinker (Stratagene). Microarrays are washed at room temperature once in 0.2% SDS and three times in distilled water. Non-specific binding sites are blocked by incubation of microarrays in 0.2% casein in phosphate buffered saline (PBS) (Tropix, Inc., Bedford MA) for 30 minutes at 60°C followed by washes in 0.2% SDS and distilled water as before.

Hybridization

Hybridization reactions contain 9 μl of sample mixture consisting of 0.2 μg each of Cy3 and Cy5 labeled cDNA synthesis products in 5X SSC, 0.2% SDS hybridization buffer. The sample mixture is heated to 65°C for 5 minutes and is aliquoted onto the microarray surface and covered with an 1.8 cm^2 coverslip. The arrays are transferred to a waterproof chamber having a cavity just slightly larger than a microscope slide. The chamber is kept at 100% humidity internally by the addition of 140 μl of 5X SSC in a corner of the chamber. The chamber containing the arrays is incubated for about 6.5 hours at 60°C. The arrays are washed for 10 min at 45°C in a first wash buffer (1X SSC, 0.1% SDS), three times for 10 minutes each at 45°C in a second wash buffer (0.1X SSC), and dried.

Detection

Reporter-labeled hybridization complexes are detected with a microscope equipped with an Innova 70 mixed gas 10 W laser (Coherent, Inc., Santa Clara CA) capable of generating spectral lines at 488 nm for excitation of Cy3 and at 632 nm for excitation of Cy5. The excitation laser light is focused on the array using a 20X microscope objective (Nikon, Inc., Melville NY). The slide containing the array is placed on a computer-controlled X-Y stage on the microscope and raster-scanned past the objective. The 1.8 cm x 1.8 cm array used in the present example is scanned with a resolution of 20 micrometers.

In two separate scans, a mixed gas multiline laser excites the two fluorophores sequentially. Emitted light is split, based on wavelength, into two photomultiplier tube detectors (PMT R1477, Hamamatsu Photonics Systems, Bridgewater NJ) corresponding to the two fluorophores. Appropriate filters positioned between the array and the photomultiplier tubes are used to filter the signals. The emission maxima of the fluorophores used are 565 nm for Cy3 and 650 nm for Cy5. Each array is typically scanned twice, one scan per fluorophore using the appropriate filters at the laser source,

although the apparatus is capable of recording the spectra from both fluorophores simultaneously.

The sensitivity of the scans is typically calibrated using the signal intensity generated by a cDNA control species added to the sample mixture at a known concentration. A specific location on the array contains a complementary DNA sequence, allowing the intensity of the signal at that location to be correlated with a weight ratio of hybridizing species of 1:100,000. When two samples from different sources (e.g., representing test and control cells), each labeled with a different fluorophore, are hybridized to a single array for the purpose of identifying genes that are differentially expressed, the calibration is done by labeling samples of the calibrating cDNA with the two fluorophores and adding identical amounts of each to the hybridization mixture.

The output of the photomultiplier tube is digitized using a 12-bit RTI-835H analog-to-digital (A/D) conversion board (Analog Devices, Inc., Norwood MA) installed in an IBM-compatible PC computer. The digitized data are displayed as an image where the signal intensity is mapped using a linear 20-color transformation to a pseudocolor scale ranging from blue (low signal) to red (high signal). The data is also analyzed quantitatively. Where two different fluorophores are excited and measured simultaneously, the data are first corrected for optical crosstalk (due to overlapping emission spectra) between the fluorophores using each fluorophore's emission spectrum.

A grid is superimposed over the fluorescence signal image such that the signal from each spot is centered in each element of the grid. The fluorescence signal within each element is then integrated to obtain a numerical value corresponding to the average intensity of the signal. The software used for signal analysis is the GEMTOOLS gene expression analysis program (Incyte).

XI. Complementary Polynucleotides

Sequences complementary to the TRICH-encoding sequences, or any parts thereof, are used to detect, decrease, or inhibit expression of naturally occurring TRICH. Although use of oligonucleotides comprising from about 15 to 30 base pairs is described, essentially the same procedure is used with smaller or with larger sequence fragments. Appropriate oligonucleotides are designed using OLIGO 4.06 software (National Biosciences) and the coding sequence of TRICH. To inhibit transcription, a complementary oligonucleotide is designed from the most unique 5' sequence and used to prevent promoter binding to the coding sequence. To inhibit translation, a complementary oligonucleotide is designed to prevent ribosomal binding to the TRICH-encoding transcript.

XII. Expression of TRICH

Expression and purification of TRICH is achieved using bacterial or virus-based expression systems. For expression of TRICH in bacteria, cDNA is subcloned into an appropriate vector containing an antibiotic resistance gene and an inducible promoter that directs high levels of cDNA transcription. Examples of such promoters include, but are not limited to, the *trp-lac* (*tac*) hybrid

promoter and the T5 or T7 bacteriophage promoter in conjunction with the *lac* operator regulatory element. Recombinant vectors are transformed into suitable bacterial hosts, e.g., BL21(DE3). Antibiotic resistant bacteria express TRICH upon induction with isopropyl beta-D-thiogalactopyranoside (IPTG). Expression of TRICH in eukaryotic cells is achieved by infecting insect or mammalian cell lines with recombinant Autographica californica nuclear polyhedrosis virus (AcMNPV), commonly known as baculovirus. The nonessential polyhedrin gene of baculovirus is replaced with cDNA encoding TRICH by either homologous recombination or bacterial-mediated transposition involving transfer plasmid intermediates. Viral infectivity is maintained and the strong polyhedrin promoter drives high levels of cDNA transcription. Recombinant baculovirus is used to infect Spodoptera frugiperda (Sf9) insect cells in most cases, or human hepatocytes, in some cases. Infection of the latter requires additional genetic modifications to baculovirus. (See Engelhard, E.K. et al. (1994) Proc. Natl. Acad. Sci. USA 91:3224-3227; Sandig, V. et al. (1996) Hum. Gene Ther. 7:1937-1945.)

In most expression systems, TRICH is synthesized as a fusion protein with, e.g., glutathione S-transferase (GST) or a peptide epitope tag, such as FLAG or 6-His, permitting rapid, single-step, affinity-based purification of recombinant fusion protein from crude cell lysates. GST, a 26-kilodalton enzyme from Schistosoma japonicum, enables the purification of fusion proteins on immobilized glutathione under conditions that maintain protein activity and antigenicity (Amersham Pharmacia Biotech). Following purification, the GST moiety can be proteolytically cleaved from TRICH at specifically engineered sites. FLAG, an 8-amino acid peptide, enables immunoaffinity purification using commercially available monoclonal and polyclonal anti-FLAG antibodies (Eastman Kodak). 6-His, a stretch of six consecutive histidine residues, enables purification on metal-chelate resins (QIAGEN). Methods for protein expression and purification are discussed in Ausubel (1995, supra, ch. 10 and 16). Purified TRICH obtained by these methods can be used directly in the assays shown in Examples XVI, XVII, and XVIII, where applicable.

XIII. Functional Assays

TRICH function is assessed by expressing the sequences encoding TRICH at physiologically elevated levels in mammalian cell culture systems. cDNA is subcloned into a mammalian expression vector containing a strong promoter that drives high levels of cDNA expression. Vectors of choice include PCMV SPORT (Life Technologies) and PCR3.1 (Invitrogen, Carlsbad CA), both of which contain the cytomegalovirus promoter. 5-10 μg of recombinant vector are transiently transfected into a human cell line, for example, an endothelial or hematopoietic cell line, using either liposome formulations or electroporation. 1-2 μg of an additional plasmid containing sequences encoding a marker protein are co-transfected. Expression of a marker protein provides a means to distinguish transfected cells from nontransfected cells and is a reliable predictor of cDNA expression from the

recombinant vector. Marker proteins of choice include, e.g., Green Fluorescent Protein (GFP; Clontech), CD64, or a CD64-GFP fusion protein. Flow cytometry (FCM), an automated, laser optics-based technique, is used to identify transfected cells expressing GFP or CD64-GFP and to evaluate the apoptotic state of the cells and other cellular properties. FCM detects and quantifies the uptake of fluorescent molecules that diagnose events preceding or coincident with cell death. These events include changes in nuclear DNA content as measured by staining of DNA with propidium iodide; changes in cell size and granularity as measured by forward light scatter and 90 degree side light scatter; down-regulation of DNA synthesis as measured by decrease in bromodeoxyuridine uptake; alterations in expression of cell surface and intracellular proteins as measured by reactivity with specific antibodies; and alterations in plasma membrane composition as measured by the binding of fluorescein-conjugated Annexin V protein to the cell surface. Methods in flow cytometry are discussed in Ormerod, M.G. (1994) Flow Cytometry, Oxford, New York NY.

The influence of TRICH on gene expression can be assessed using highly purified populations of cells transfected with sequences encoding TRICH and either CD64 or CD64-GFP. CD64 and CD64-GFP are expressed on the surface of transfected cells and bind to conserved regions of human immunoglobulin G (IgG). Transfected cells are efficiently separated from nontransfected cells using magnetic beads coated with either human IgG or antibody against CD64 (DYNAL, Lake Success NY). mRNA can be purified from the cells using methods well known by those of skill in the art. Expression of mRNA encoding TRICH and other genes of interest can be analyzed by northern analysis or microarray techniques.

XIV. Production of TRICH Specific Antibodies

TRICH substantially purified using polyacrylamide gel electrophoresis (PAGE; see, e.g., Harrington, M.G. (1990) Methods Enzymol. 182:488-495), or other purification techniques, is used to immunize rabbits and to produce antibodies using standard protocols.

Alternatively, the TRICH amino acid sequence is analyzed using LASERGENE software (DNASTAR) to determine regions of high immunogenicity, and a corresponding oligopeptide is synthesized and used to raise antibodies by means known to those of skill in the art. Methods for selection of appropriate epitopes, such as those near the C-terminus or in hydrophilic regions are well described in the art. (See, e.g., Ausubel, 1995, supra, ch. 11.)

Typically, oligopeptides of about 15 residues in length are synthesized using an ABI 431A peptide synthesizer (Applied Biosystems) using FMOC chemistry and coupled to KLH (Sigma-Aldrich, St. Louis MO) by reaction with N-maleimidobenzoyl-N-hydroxysuccinimide ester (MBS) to increase immunogenicity. (See, e.g., Ausubel, 1995, supra.) Rabbits are immunized with the oligopeptide-KLH complex in complete Freund's adjuvant. Resulting antisera are tested for antipeptide and anti-TRICH activity by, for example, binding the peptide or TRICH to a substrate,

blocking with 1% BSA, reacting with rabbit antisera, washing, and reacting with radio-iodinated goat anti-rabbit IgG.

XV. Purification of Naturally Occurring TRICH Using Specific Antibodies

Naturally occurring or recombinant TRICH is substantially purified by immunoaffinity chromatography using antibodies specific for TRICH. An immunoaffinity column is constructed by covalently coupling anti-TRICH antibody to an activated chromatographic resin, such as CNBr-activated SEPHAROSE (Amersham Pharmacia Biotech). After the coupling, the resin is blocked and washed according to the manufacturer's instructions.

Media containing TRICH are passed over the immunoaffinity column, and the column is washed under conditions that allow the preferential absorbance of TRICH (e.g., high ionic strength buffers in the presence of detergent). The column is eluted under conditions that disrupt antibody/TRICH binding (e.g., a buffer of pH 2 to pH 3, or a high concentration of a chaotrope, such as urea or thiocyanate ion), and TRICH is collected.

XVI. Identification of Molecules Which Interact with TRICH

Molecules which interact with TRICH may include transporter substrates, agonists or antagonists, modulatory proteins such as $G\beta\gamma$ proteins (Reimann, supra) or proteins involved in TRICH localization or clustering such as MAGUKs (Craven, supra). TRICH, or biologically active fragments thereof, are labeled with ^{125}I Bolton-Hunter reagent. (See, e.g., Bolton A.E. and W.M. Hunter (1973) Biochem. J. 133:529-539.) Candidate molecules previously arrayed in the wells of a multi-well plate are incubated with the labeled TRICH, washed, and any wells with labeled TRICH complex are assayed. Data obtained using different concentrations of TRICH are used to calculate values for the number, affinity, and association of TRICH with the candidate molecules.

Alternatively, proteins that interact with TRICH are isolated using the yeast 2-hybrid system (Fields, S. and O. Song (1989) Nature 340:245-246). TRICH, or fragments thereof, are expressed as fusion proteins with the DNA binding domain of Gal4 or lexA, and potential interacting proteins are expressed as fusion proteins with an activation domain. Interactions between the TRICH fusion protein and the TRICH interacting proteins (fusion proteins with an activation domain) reconstitute a transactivation function that is observed by expression of a reporter gene. Yeast 2-hybrid systems are commercially available, and methods for use of the yeast 2-hybrid system with ion channel proteins are discussed in Niethammer, M. and M. Sheng (1998, Meth. Enzymol. 293:104-122).

TRICH may also be used in the PATHCALLING process (CuraGen Corp., New Haven CT) which employs the yeast two-hybrid system in a high-throughput manner to determine all interactions between the proteins encoded by two large libraries of genes (Nandabalan, K. et al. (2000) U.S. Patent No. 6,057,101).

Potential TRICH agonists or antagonists may be tested for activation or inhibition of TRICH

ion channel activity using the assays described in section XVIII.

XVII. Demonstration of TRICH Activity

Ion channel activity of TRICH is demonstrated using an electrophysiological assay for ion conductance. TRICH can be expressed by transforming a mammalian cell line such as COS7, HeLa or CHO with a eukaryotic expression vector encoding TRICH. Eukaryotic expression vectors are commercially available, and the techniques to introduce them into cells are well known to those skilled in the art. A second plasmid which expresses any one of a number of marker genes, such as ß-galactosidase, is co-transformed into the cells to allow rapid identification of those cells which have taken up and expressed the foreign DNA. The cells are incubated for 48-72 hours after transformation under conditions appropriate for the cell line to allow expression and accumulation of TRICH and ß-galactosidase.

Transformed cells expressing ß-galactosidase are stained blue when a suitable colorimetric substrate is added to the culture media under conditions that are well known in the art. Stained cells are tested for differences in membrane conductance by electrophysiological techniques that are well known in the art. Untransformed cells, and/or cells transformed with either vector sequences alone or ß-galactosidase sequences alone, are used as controls and tested in parallel. Cells expressing TRICH will have higher anion or cation conductance relative to control cells. The contribution of TRICH to conductance can be confirmed by incubating the cells using antibodies specific for TRICH. The antibodies will bind to the extracellular side of TRICH, thereby blocking the pore in the ion channel, and the associated conductance.

Alternatively, ion channel activity of TRICH is measured as current flow across a TRICH-containing Xenopus laevis oocyte membrane using the two-electrode voltage-clamp technique (Ishi et al., supra; Jegla, T. and L. Salkoff (1997) J. Neurosci. 17:32-44). TRICH is subcloned into an appropriate Xenopus oocyte expression vector, such as pBF, and 0.5-5 ng of mRNA is injected into mature stage IV oocytes. Injected oocytes are incubated at 18 °C for 1-5 days. Inside-out macropatches are excised into an intracellular solution containing 116 mM K-gluconate, 4 mM KCl, and 10 mM Hepes (pH 7.2). The intracellular solution is supplemented with varying concentrations of the TRICH mediator, such as cAMP, cGMP, or Ca^{+2} (in the form of $CaCl_2$), where appropriate. Electrode resistance is set at 2-5 MΩ and electrodes are filled with the intracellular solution lacking mediator. Experiments are performed at room temperature from a holding potential of 0 mV. Voltage ramps (2.5 s) from -100 to 100 mV are acquired at a sampling frequency of 500 Hz. Current measured is proportional to the activity of TRICH in the assay.

In particular, the activity of TRICH-20 is measured as Ca^{2+} conductance, the activity of TRICH-22 is measured as Cl^- conductance in the presence of glycine, the activity of TRICH-23 is

measured as Ca^{2+} conductance, and the activity of TRICH-24 is measured as K^+ conductance in the presence of Ca^{2+}, and the activity of TRICH-26 is measured as cation conductance in the presence of heat.

Transport activity of TRICH is assayed by measuring uptake of labeled substrates substrates (including but not limited to, maltose, glucose, or glycogen) into Xenopus laevis oocytes. Oocytes at stages V and VI are injected with TRICH mRNA (10 ng per oocyte) and incubated for 3 days at 18°C in OR2 medium (82.5mM NaCl, 2.5 mM KCl, 1mM $CaCl_2$, 1mM $MgCl_2$, 1mM Na_2HPO_4, 5 mM Hepes, 3.8 mM NaOH, 50μg/ml gentamycin, pH 7.8) to allow expression of TRICH. Oocytes are then transferred to standard uptake medium (100mM NaCl, 2 mM KCl, 1mM $CaCl_2$, 1mM $MgCl_2$, 10 mM Hepes/Tris pH 7.5). Uptake of various substrates (e.g., amino acids, sugars, drugs, ions, and neurotransmitters) is initiated by adding labeled substrate (e.g. radiolabeled with ^3H, fluorescently labeled with rhodamine, etc.) to the oocytes. After incubating for 30 minutes, uptake is terminated by washing the oocytes three times in Na^+-free medium, measuring the incorporated label, and comparing with controls. TRICH activity is proportional to the level of internalized labeled substrate. In particular, test substrates include sulfate for TRICH-13, tricarboxylates for TRICH-21, dicarboxylates and Na^+ for TRICH-25, ornithine for TRICH-27, and monocarboxylates for TRICH-28.

ATPase activity associated with TRICH can be measured by hydrolysis of radiolabeled ATP-[γ-^{32}P], separation of the hydrolysis products by chromatographic methods, and quantitation of the recovered ^{32}P using a scintillation counter. The reaction mixture contains ATP-[γ-^{32}P] and varying amounts of TRICH in a suitable buffer incubated at 37°C for a suitable period of time. The reaction is terminated by acid precipitation with trichloroacetic acid and then neutralized with base, and an aliquot of the reaction mixture is subjected to membrane or filter paper-based chromatography to separate the reaction products. The amount of ^{32}P liberated is counted in a scintillation counter. The amount of radioactivity recovered is proportional to the ATPase activity of TRICH in the assay.

XVIII. Identification of TRICH Agonists and Antagonists

TRICH is expressed in a eukaryotic cell line such as CHO (Chinese Hamster Ovary) or HEK (Human Embryonic Kidney) 293. Ion channel activity of the transformed cells is measured in the presence and absence of candidate agonists or antagonists. Ion channel activity is assayed using patch clamp methods well known in the art or as described in Example XVII. Alternatively, ion channel activity is assayed using fluorescent techniques that measure ion flux across the cell membrane (Velicelebi, G. et al. (1999) Meth. Enzymol. 294:20-47; West, M.R. and C.R. Molloy (1996) Anal. Biochem. 241:51-58). These assays may be adapted for high-throughput screening using microplates. Changes in internal ion concentration are measured using fluorescent dyes such as the Ca^{2+} indicator Fluo-4 AM, sodium-sensitive dyes such as SBFI and sodium green, or the Cl^-

indicator MQAE (all available from Molecular Probes) in combination with the FLIPR fluorimetric plate reading system (Molecular Devices). In a more generic version of this assay, changes in membrane potential caused by ionic flux across the plasma membrane are measured using oxonyl dyes such as $DiBAC_4$ (Molecular Probes). $DiBAC_4$ equilibrates between the extracellular solution and cellular sites according to the cellular membrane potential. The dye's fluorescence intensity is 20-fold greater when bound to hydrophobic intracellular sites, allowing detection of $DiBAC_4$ entry into the cell (Gonzalez, J.E. and P.A. Negulescu (1998) Curr. Opin. Biotechnol. 9:624-631). Candidate agonists or antagonists may be selected from known ion channel agonists or antagonists, peptide libraries, or combinatorial chemical libraries.

Various modifications and variations of the described methods and systems of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention. Although the invention has been described in connection with certain embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments. Indeed, various modifications of the described modes for carrying out the invention which are obvious to those skilled in molecular biology or related fields are intended to be within the scope of the following claims.

Table 1

Incyte Project ID	Polypeptide SEQ ID NO:	Incyte Polypeptide ID	Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID
2194064	1	2194064CD1	31	2194064CB1
2744094	2	2744094CD1	32	2744094CB1
2798241	3	2798241CD1	33	2798241CB1
3105257	4	3105257CD1	34	3105257CB1
3200979	5	3200979CD1	35	3200979CB1
6754139	6	6754139CD1	36	6754139CB1
6996659	7	6996659CD1	37	6996659CB1
7472747	8	7472747CD1	38	7472747CB1
7474121	9	7474121CD1	39	7474121CB1
7475615	10	7475615CD1	40	7475615CB1
7475656	11	7475656CD1	41	7475656CB1
7480632	12	7480632CD1	42	7480632CB1
6952742	13	6952742CD1	43	6952742CB1
7478795	14	7478795CD1	44	7478795CB1
656293	15	656293CD1	45	656293CB1
7473957	16	7473957CD1	46	7473957CB1
7474111	17	7474111CD1	47	7474111CB1
7480826	18	7480826CD1	48	7480826CB1
6025572	19	6025572CD1	49	6025572CB1
5686561	20	5686561CD1	50	5686561CB1
1553725	21	1553725CD1	51	1553725CB1
1695770	22	1695770CD1	52	1695770CB1
4672222	23	4672222CD1	53	4672222CB1
6176128	24	6176128CD1	54	6176128CB1
7473418	25	7473418CD1	55	7473418CB1
7474129	26	7474129CD1	56	7474129CB1
7481414	27	7481414CD1	57	7481414CB1
7481461	28	7481461CD1	58	7481461CB1
7472541	29	7472541CD1	59	7472541CB1
6999183	30	6999183CD1	60	6999183CB1

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO:	Probability score	GenBank Homolog
1	2194064CD1	g2463634	1.60E-41	Monocarboxylate transporter [Homo sapiens] (Price, N. T. et al. (1998) Biochem. J. 329:321-328)
2	2744094CD1	g13346481	0	ATP-binding cassette transporter MRP8 [Homo sapiens]
3	2798241CD1	g1699038	2.90E-142	ABC3 [Homo sapiens] (Connors, T. D. et al. (1997) Genomics 39:231-234)
4	3105257CD1	g8650412	0	M-ABC2 protein [Homo sapiens] (Zhang, F. et al. (2000) Characterization of ABCB9, an ATP binding cassette protein associated with lysosomes J. Biol. Chem. 275:23287-23294)
5	3200979CD1	g1514530	3.10E-119	ABC-C transporter [Homo sapiens] (Klugbauer, N. and F. Hofmann (1996) FEBS Lett. 391:61-65)
6	6754139CD1	g6746563	1.70E-188	neuronal nicotinic acetylcholine receptor subunit [Rattus norvegicus] (Elgoyhen, A. B. et al. (2001) alpha 10: A determinant of nicotinic cholinergic receptor function in mammalian vestibular and cochlear mechanosensory hair cells Proc. Natl. Acad. Sci. U.S.A. 98:3501-3506)
7	6996659CD1	g1050330	0	Ionotropic glutamate receptor [Rattus norvegicus] (Ciabarra, A.M. et al. (1995) J. Neurosci. 15:6498-6508)
8	7472747CD1	g13926108	1.00E-157	2P domain potassium channel Talk-1 [Homo sapiens] (Girard,C. et al. (2001) Genomic and functional characteristics of novel human pancreatic 2P domain K(+) channels. Biochem Biophys Res Commun. 282:249-256)
9	7474121CD1	g2465542	7.00E-20	TWIK-related acid-sensitive K+ channel [Homo sapiens] (Duprat, F. et al. (1997) EMBO J. 16:5464-5471)
10	7475615CD1	g2654005	5.70E-114	Pendrin [Homo sapiens] (Everett, L.A. et al. (1997) Nature Genet. 17:411-422)

Table 2 (cont.)

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO:	Probability score	GenBank Homolog
11	7475656CD1	g3168874	0	Ion channel BCNG-1 [Homo sapiens] (Santoro, B. et al. (1997) Proc. Natl. Acad. Sci. USA 94:14815-14820)
12	7480632CD1	g1514530	9.80E-123	ABC-C transporter [Homo sapiens] (Klugbauer, N. and F. Hofmann (1996) FEBS Lett. 391:61-65)
13	6952742CD1	g10719650	0	sulfate/anion transporter SAT-1 protein [Homo sapiens] (Lohi,H. et al. (2000) Mapping of Five New Putative Anion Transporter Genes in Human and Characterization of SLC26A6, A Candidate Gene for Pancreatic Anion Exchanger. Genomics 70:102-112)
		g431453	3.10E-276	Sulfate anion transporter [Rattus norvegicus] (Bissig,M. et al. (1994) Functional expression cloning of the canalicular sulfate transport system of rat hepatocytes. J. Biol. Chem. 269:3017-3021)
14	7478795CD1	g6045150	0	TAP-like ABC transporter [Rattus norvegicus] (Yamaguchi,Y. et al. (1999) An ABC transporter homologous to TAP proteins. FEBS Lett. 457:231-236)
15	656293CD1	g6746563	1.30E-220	neuronal nicotinic acetylcholine receptor [Rattus norvegicus]
16	7473957CD1	g340199	1.20E-130	voltage-dependent anion channel [Homo sapiens] (Blachly-Dyson, E. et al. (1993) J. Biol. Chem. 268:1835-1841)
17	7474111CD1	g6006493	1.50E-75	Cardiac potassium channel subunit (Kv6.2) [Homo sapiens] (Zhu, X., et al. (1999) Receptors Channels 6:337-350)
18	7480826CD1	g8248427	1.50E-235	amino acid transporter system A [Rattus norvegicus] (Sugawara, M. et al. (2000) J. Biol. Chem. 275:16473-16477)
19	6025572CD1	g402628	4.20E-114	adenine nucleotide carrier [Mus musculus]

Table 2 (cont.)

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO:	Probability score	GenBank Homolog
20	5686561CD1	g4586963	2.40E-27	voltage-gated ca channel [Rattus norvegicus] (Ishibashi, K. et al. (2000) Molecular cloning of a novel form (Two-repeat) protein related to voltage-gated sodium and calcium channels. Biochem. Biophys. Res. Commun. 270:370-376)
21	1553725CD1	g545998	1.60E-89	tricarboxylate carrier [Rattus sp.] (Azzi, A. et al. (1993) The mitochondrial tricarboxylate carrier. J. Bioenerg. Biomembr. 25:515-524)
22	1695770CD1	g31849	1.10E-175	inhibitory glycine receptor [Homo sapiens] (Grenningloh, G. et al. (1990) Alpha subunit variants of the human glycine receptor: primary structures, functional expression and chromosomal localization of the corresponding genes. EMBO J. 9:771-776)
23	4672222CD1	g13562153	0	channel-kinase 1 [Homo sapiens] (Ryazanov, A. G. et al. (1999) Alpha-kinases: a new class of protein kinases with a novel catalytic domain Curr. Biol. 9:R43-R45)
24	6176128CD1	g3978472	0	potassium channel subunit [Rattus norvegicus] (Joiner, W.J. et al. (1998) Formation of intermediate-conductance calcium-activated potassium channels by interaction of Slack and Slo subunits. Nat Neurosci. 1:462-469)
25	7473418CD1	g2811122	2.90E-177	NaDC-2 [Xenopus laevis]
26	7474129CD1	g2570933	1.20E-134	vanilloid receptor subtype 1 [Rattus norvegicus] (Caterina, M.J. et al. (1997) The capsaicin receptor: a heat-activated ion channel in the pain pathway. Nature 389:816-824)

Table 2 (cont.)

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO:	Probability score	GenBank Homolog
27	7481414CD1	g13445630	1.00E-151	mutant ornithine transporter 2 [Mus musculus] (Wu, Q. and Maniatis, T. (1999) A striking organization of a large family of human neural cadherin-like cell adhesion genes. Cell 97:779-790)
28	7481461CD1	g458247	1.40E-136	X-linked PEST-containing transporter [Homo sapiens] (Lafreniere, R.G. et al. (1994) A novel transmembrane transporter encoded by the XPCT gene in Xq13.2. Mol. Genet. 3:1133-1139)
29	7472541CD1	g6457270	0	Putative E1-E2 ATPase [Mus musculus] (Halleck, M.S. et al. (1999) Differential expression of putative transbilayer amphipath transporters. Physiol. Genomics (Online) 1:139-150)
30	6999183CD1	g1514530	2.30E-127	ABC-C transporter [Homo sapiens] (Klugbauer N. and Hofmann F.(1996) Primary structure of a novel ABC transporter with a chromosomal localization on the band encoding the multidrug resistance-associated protein, FEBS Lett. 391:61-65)

Table 3

<table>
<tr><th>SEQ ID NO:</th><th>Incyte Polypeptide ID</th><th>Amino Acid Residues</th><th>Potential Phosphorylation Sites</th><th>Potential Glycosylation Sites</th><th>Signature Sequences, Domains and Motifs</th><th>Analytical Methods and Databases</th></tr>
<tr><td rowspan="3">1</td><td rowspan="3">2194064CD1</td><td rowspan="3">308</td><td rowspan="3">S287 S51 T132</td><td rowspan="3"></td><td>Signal peptide:
M1-A17</td><td>SPScan</td></tr>
<tr><td>Transmembrane domains:
W197-V224, Y248-G270</td><td>HMMER</td></tr>
<tr><td>PEST transporter:
DM05037|P53988|1-465:M1-L109, L126-K289
DM05037|Q03064|1-475:M1-L109, V110-K289
DM05037|P36021|155-612:G3-G288</td><td>BLAST-DOMO</td></tr>
<tr><td rowspan="10">2</td><td rowspan="10">2744094CD1</td><td rowspan="10">606</td><td rowspan="10">S116 S133 S266 S299 S403 S503 S604 S63 T112 T253 T318 T330 T388 T455 T543 T70</td><td rowspan="10">N216 N386 N62 N68</td><td>Transmembrane domains:
P25-W49, Q82-I107, L166-L187, P184-M203</td><td>HMMER</td></tr>
<tr><td>ABC transporter:
H392-G575</td><td>HMMER-PFAM</td></tr>
<tr><td>ABC transporter transmembrane region:
S30-A319</td><td>HMMER-PFAM</td></tr>
<tr><td>ABC transporters family signature:
A483-D533</td><td>ProfileScan</td></tr>
<tr><td>ABC transporter:
F502-V516</td><td>MOTIFS</td></tr>
<tr><td>ATP/GTP binding site:
G399-S406</td><td>MOTIFS</td></tr>
<tr><td>ATP-binding transporter:
PD00131:G141-D150, S403-I456, G550-R587</td><td>BLIMPS-PRODOM</td></tr>
<tr><td>ABC transporters family:
DM00008|P33527|1293-1502: F367-G575
DM00008|Q10185|1239-1448: I365-G575
DM00008|P39109|1272-1482: I365-G575
DM00008|S64757|1302-1528: I365-K486</td><td>BLAST-DOMO</td></tr>
<tr><td>ATP-binding transport protein:
PD000130: T61-G292
PD002040: G434-P488</td><td>BLAST-PRODOM</td></tr>
</table>

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
3	2798241CD1	1642	S199 S32 T1366 S431 S443 S1367 S460 S546 T1390 S582 S616 S1405 S624 S761 T1454 S815 S859 T1461 S861 S885 S1558 S962 T28 T1635 T486 T518 T1099 T572 T606 T1126 T779 T780 S1190 T854 Y168 S1236 S1247 S1308 S1372 T1429 Y1552	N190 N388 N458 N499 N576 N86 N943 N973 N996 N1245 N1556 N1627	Transmembrane domains: Q34-M52, S272-P292, S295-F313, V327-I346, I401-L427, V865-H883, P1075-Y1098, L1095-P1114, W1137-I1162, I1165-I1184	HMMER
					ABC transporter: G507-G689, G1326-G1509	HMMER-PFAM
					ABC transporters family signature: V595D646, I1413-D1464	ProfileScan
					ABC transporter: L615-V629	MOTIFS
					ATP/GTP binding sites: G514-S521, G1333-S1340	MOTIFS
					ABC transporters family: DM00008|P41233|839-1045:I478-S687, K1313-M1506	BLAST-DOMO
					ABC transporters family: DM00008|P34358|611-816:I478-S687, I1319-M1506 DM00008|P26050|8-212:K1313-S1508, I478-I686 DM00008|P41233|1851-2058:R1309-S1508, I478-I686	BLAST-DOMO

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
4	3105257CD1	659	S206 S26 S300 S452 S504 S583 S62 T261 T284 T293 T348 T520 T615 Y121 Y298	N131 N210	ABC transporter: G441-G628	HMMER-PFAM
					ABC transporter transmembrane region: L92-I366	HMMER-PFAM
					ABC transporters family signature: A535-D586	ProfileScan
					ABC transporter: L555-L569	MOTIFS
					ATP/GTP binding site: G448-S455	MOTIFS
					ABC transporters family: BL00211: L446-V457, L555-D586	BLIMPS-BLOCKS
					ATP-binding transporter: PD00131: G190-D199, S452-I505, G603-L640	BLIMPS-PRODOM
					ABC transporters family: DM00008|A42150|367-576: L413-L625 DM00008|P34712|1076-1290: F415-G628	BLAST-DOMO
					ATP-binding transport protein: PD000130: L135-Y358	BLAST-PRODOM
					Multidrug resistance ATP-binding transport protein: PD167072: W486-G552	BLAST-PRODOM

Table 3 (cont.)

<table>
<tr><th>SEQ ID NO:</th><th>Incyte Polypeptide ID</th><th>Amino Acid Residues</th><th>Potential Phosphorylation Sites</th><th>Potential Glycosylation Sites</th><th>Signature Sequences, Domains and Motifs</th><th>Analytical Methods and Databases</th></tr>
<tr><td rowspan="6">5</td><td rowspan="6">3200979CD1</td><td rowspan="6">1592</td><td rowspan="6">S125 S187 T1117 S207 S386 T1135 S453 Y906 T1214 S714 S733 T1346 S745 S770 T1388 S778 S874 T1417 S882 S994 S1454 T368 T439 T1494 T484 T542 T1580 T565 T673 S1116 T691 T706 S1206 T766 T1257 T782 T801 T1264 T927 T98 T1265 Y1192 S7 S1297 S1320 T77 S1328 T1434 T1466</td><td rowspan="6">N185 N62 N75 N870 N871 N899 N949 N1164 N1273</td><td>Transmembrane domains:
I265-V285, L296-I315, M319-L340, I390-F410, L815-M834, L1063-M1082, W1099-T1117, L1126-L1145</td><td>HMMER</td></tr>
<tr><td>ABC transporter:
G500-G642, G1281-G1465</td><td>HMMER-PFAM</td></tr>
<tr><td>ABC transporters family signature:
L1372-D1420</td><td>ProfileScan</td></tr>
<tr><td>ATP/GTP binding sites:
G507-S514, G1288-S1295</td><td>MOTIFS</td></tr>
<tr><td>ABC transporters family:
BL00211: I505-L516, L1389-D1420</td><td>BLIMPS-BLOCKS</td></tr>
<tr><td>ABC transporters family:
DM00008|P41233|839-1045:K1268-M1462, I471-P600, E587-N641
DM00008|P34358|611-816:F1262-M1462, I471-D592, E585-N641
DM00008|P41233|1851-2058:K1266-S1464, I471-V584, V588-N641
DM00008|P23703|41-246:K1268-G1465, V476-L609, E585-G642</td><td>BLAST-DOMO</td></tr>
</table>

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
6	6754139CD1	382	S124 S260 S340 S85 T337		Transmembrane domains: A168-H191, V200-L217, Y233-N253, F361-L378	HMMER
					Neurotransmitter-gated ion channel: D2-L378	HMMER-PFAM
					Neurotransmitter-gated ion-channels signature: V66-G120	ProfileScan
					Neurotransmitter-gated ion channel: C86-C100	MOTIFS
					Neurotransmitter-gated ion channel: BL00236:M1-D26, Y155-S196, V43-N52, D71-H109	BLIMPS-BLOCKS
					Neurotransmitter-gated ion channel: PR00252:T9-W25, L42-K53, C86-C100, L162-N174	BLIMPS-PRINTS
					Nicotinic acetylcholine channel: PR00254:M1-L12, Y30-W44, I48-G60, V66-S84	BLIMPS-PRINTS
					Neurotransmitter-gated ion channel: DM00195|P43144|5-478:M1-E296, R323-A381 DM00195|JH0173|14-503:M1-P314, L327-A381 DM00195|P09478|5-538:R4-L297, E296-A381 DM00195|P54131|3-491:M1-A312, L327-A381	BLAST-DOMO
					Postsynaptic ion channel: PD000153: M1-R262, S298-V377	BLAST-PRODOM

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
7	6996659CD1	1115	S110 S202 S1030 S248 S303 S1080 S317 S334 T1101 S383 S448 S1098 S5 S552 T1109 S563 S800 S801 S809 S976 S986 T441 T519 T636 T693 T704 T741 T796 T85 T949 T997 Y1106	N145 N264 N275 N285 N296 N426 N439 N549 N565 N709 N886 N965 N984 N1015 N1018 N1069	Signal peptide: M1-V24	HMMER
					Signal peptide: M1-S33	SPScan
					Transmembrane domains: M677-T693, F931-I946	HMMER
					Ligand-gated ion channel: H674-E952	HMMER-PFAM
					ATP/GTP binding site: G373-T380	MOTIFS
					NMDA receptor signature: PR00177:M677-G702, F744-E771,F931-V955, F593-L621	BLIMPS-PRINTS
					Glutamate receptor: DM00247|P35436|615-886: T731-Q993 DM00247|Q03391|640-919: T731-Y956 DM00393|Q01097|377-614: G482-F728 DM00247|Q01097|616-887: T731-Y956	BLAST-DOMO
					Ionotropic glutamate receptor: PD156309: S170-Y577 PD139812: M1-P169 PD124284: S986-S1115 PD000500: M670-E952	BLAST-PRODOM
8	7472747CD1	295	S193 S199 S91 T59	N57 N86	Signal peptide: M1-A41	SPScan
					Transmembrane domains: F95-L114, V167-F187	HMMER
9	7474121CD1	384	S205 S252 S267 S42 T306 T329 T74	N70 N96	Transmembrane domains: G23-A43, F103-I122, L132-D150, F337-Q359	HMMER

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
10	7475615CD1	769	S200 S3 S407 S461 S475 S572 S651 S707 S738 S742 S748 S87 T15 T282 T60 Y470 Y57	N195 N198 N596	Transmembrane domains: F245-I265, N294-V311, F491-V510	HMMER
					Sulfate transporter family: L229-T513	HMMER-PFAM
					Sulfate transporters profile: BL01130: G119-V172, T217-L268	BLIMPS-BLOCKS
					Sulfate transporter: DM01229|P40879|5-462: R49-V456 DM01229|P50443|49-505: E67-P495 DM01229|P45380|10-468: K78-S485 DM01229|Q02920|1-447: S87-I481	BLAST-DOMO
					Sulfate transporter protein: PD001121: V93-T197 PD001755: H641-R720, L521-D579	BLAST-PRODOM

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
11	7475656CD1	882	S102 S108 S13 S324 S360 S394 S395 S518 S544 S591 T190 T242 T649 T754 T799 T869 Y240 Y529	N330 N640 N770 N8	Transmembrane domains: L139-F159, T242-L258, I366-L392	HMMER
					Transmembrane region cyclic nucleotide domain: Y209-I453	HMMER-PFAM
					Cyclic nucleotide-binding domain: K482-M570	HMMER-PFAM
					Cyclic nucleotide-binding domain: I494-I515	MOTIFS
					Cyclic nucleotide-binding site: BL00888: G491-V514, G527-L536	BLIMPS-BLOCKS
					Cyclic nucleotide-binding domain: DM01165|A55251|333-706: H302-E576 DM01165|P29973|311-684: H302-E576 DM01165|Q03041|286-658: H302-E576 DM01165|S52072|262-635: H302-R572	BLAST-DOMO
					Cyclic nucleotide gated hyperpolarization activated cation channel: PD079330: P747-L882 PD089437: A627-M722 PD108745: M1-D62 PD151315: T577-Q626	BLAST-PRODOM

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
12	7480632CD1	1547	S134 S196 S1102 S216 S395 T1301 S7 T1343 T1389 S723 S742 T1372 S754 S779 S1405 S787 S883 T1449 S891 T107 T1535 T377 T448 S1158 T493 T551 T1212 T574 T682 S1218 T700 T715 T1219 T775 T791 S1252 T810 T86 S1275 T936 T975 S1283 Y915 S462 T1421 Y1144	N194 N71 N84 N879 N880 N908 N958 N1100 N1228	Transmembrane domains: I274-V294, L305-I324, M328-L349, I399-F419, L824-M843, M946-I963, L1021-F1040, L1046-L1064, D1105-F1123	HMMER
					ABC transporter: G509-G651, G1236-G1420	HMMER-PFAM
					ABC transporters family signature: L1327-D1375	ProfileScan
					ATP/GTP binding sites: G516-S523, G1243-S1250	MOTIFS
					ABC transporters family: BL00211: I514-L525, L1344-D1375	BLIMPS-BLOCKS
					ABC transporters family: DM00008|P41233|839-1045:K1223-M1417, I480-P609, E596-N650 DM00008|P41233|1851-2058:R1220-S1419, I480-V593, V597-N650 DM00008|P34358|611-816:F1217-M1417, I480-D601, E594-N650 DM00008|P23703|41-246:K1223-G1420, V485-L618, E594-G651	BLAST-DOMO

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
13	6952742CD1	698	S278 S355 S367 S446 S464 S594 S676 T114 T523 T559 T626 T667 T683 Y519	N155 N160	SULFATE TRANSPORTERS: DM01229|P45380|10-468: V15-R462	BLAST-DOMO
					do TRANSPORTER; SULFATE; DM08211|P45380|470-702: M463-L698	BLAST-DOMO
					PROTEIN TRANSPORT SULFATE TRANSPORTER TRANSMEMBRANE PERMEASE INTERGENIC REGION AFFINITY GLYCOPROTEIN PD001255: L285-L498	BLAST-PRODOM
					SULFATE TRANSPORTER TRANSPORT PROTEIN TRANSMEMBRANE GLYCOPROTEIN AFFINITY SULPHATE HIGH PERMEASE PD001121: L49-R136	BLAST-PRODOM
					SULFATE TRANSPORTER PROTEIN TRANSPORT TRANSMEMBRANE AFFINITY GLYCOPROTEIN SULPHATE HIGH DISEASE PD001755: H607-R689, A508-F551	BLAST-PRODOM
					SULFATE ANION TRANSPORTER 1 CANALICULAR SULFATE/CARBONATE ANTIPORTER TRANSPORT TRANSMEMBRANE GLYCOPROTEIN PD083148: D135-L191	BLAST-PRODOM
					Sulfate transporters proteins BL01130: A180-V231, D72-L125	BLIMPS-BLOCKS
					Transmembrane domain: E67-Y87, L411-A428	HMMER
					Sulfate transporter family Sulfate_transp: M192-T502	HMMER-PFAM
					Sulfate_Transporter: P95-R116	MOTIFS

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
14	7478795CD1	766	S161 S275 S28 S33 S354 S46 S543 S571 S595 S671 S702 S763 T139 T153 T181 T209 T311 T367 T377 T512 Y602	N280 N508 N524 N599 N761	MALK PROTEIN: DM00130|S13426|168-477: L195-G502	BLAST-DOMO
					ATP-BINDING TRANSPORT PROTEIN TRANSMEMBRANE GLYCOPROTEIN TRANSPORTER MULTIDRUG RESISTANCE ABC PGLYCOPROTEIN PD000130: V229-L455	BLAST-PRODOM
					ATP-BINDING TRANSPORT TRANSMEMBRANE REGION PD00131:G283-D292, S543-I596, K691-L728	BLIMPS-PRODOM
					Transmembrane domain: V85-F104, V185-F204, L328-G347, Y411-G431	HMMER
					ABC transporter transmembrane region. ABC_membrane: L188-M459	HMMER-PFAM
					ABC transporter ABC_tran: G532-G716	HMMER-PFAM
					Abc_Transporter: L643-L657	MOTIFS
					ATP/GTP-binding site motif A (P-loop) Atp_Gtp_A: G539-S546	MOTIFS
					ABC transporters family signature atp_bind_transport.prf: I625-D674	PROFILESCAN

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
15	656293CD1	450	S153 S192 S328 S408 T405	N40 N56	NEUROTRANSMITTER-GATED ION-CHANNELS DM00195|P43144|5-478:A25-E364, R391-A449	BLAST_DOMO
					CHANNEL IONIC TRANSMEMBRANE GLYCOPROTEIN POSTSYNAPTIC MEMBRANE RECEPTOR PRECURSOR SIGNAL PROTEIN PD000153:S131-R361	BLAST_PRODOM
					Neurotransmitter-gated ion channel BL00236:D139-H177, Y223-S264, V57-D94, V111-N120	BLIMPS_BLOCKS
					NEUROTRANSMITTER-GATED Ion Channel PR00252:T77-W93, L110-K121, C154-C168, L230-N242	BLIMPS_PRINTS
					NICOTINIC ACETYLCHOLINE RECEPTOR SIGNATURE PR00254:V134-S152, S64-L80, Y98-W112, I116-G128	BLIMPS_PRINTS
					signal peptide: M1-G24	HMMER
					transmembrane domain: A236-H259, V268-L285, Y301-N321, F429-L446	HMMER
					Neurotransmitter-gated ion-channel neur_chan:A30-L446	HMMER_PFAM
					Neurotr_Ion_Channel C154-C168	MOTIFS
					Neurotransmitter-gated ion-channels signature neurotr_ion_channel.prf:V134-G188	PROFILESCAN
					signal_cleavage: M1-G24	SPSCAN

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
16	7473957CD1	260	S114 S12 S211 T136 T227 T28 T47 T49 T63 T84	N215 N216	EUKARYOTIC MITOCHONDRIAL PORIN DM01893|P45879|1-282:S12-A260	BLAST_DOMO
					PORIN CHANNEL VOLTAGEDEPENDENT OUTER MEMBRANE PROTEIN MITOCHONDRION ANIONSELECTIVE MITOCHONDRIAL VDAC PD003211:A15-Q259	BLAST_PRODOM
					Eukaryotic mitochondrial porin BL00558:G33-L46, T57-S81	BLIMPS_BLOCKS
					EUKARYOTIC PORIN SIGNATURE PR00185:G45-T60, E124-E135, Y224-D241	BLIMPS_PRINTS
					Eukaryotic porin Euk_porin:A5-A260	HMMER_PFAM
					Eukaryotic_Porin Y202-Y224	MOTIFS
					Eukaryotic mitochondrial porin signature eukaryotic_porin.prf:M16-S81	PROFILESCAN
17	7474111CD1	506	S187 S194 S2 S231 S286 S423 S493 S57 T241 T273 T357 T385	N284	do CHANNEL; POTASSIUM; CDRK; FORM; DM00436|JH0595|144-307:P230-I366	BLAST_DOMO
					CHANNEL IONIC PROTEIN POTASSIUM SUBUNIT VOLTAGEGATED TRANSMEMBRANE CALCIUM TRANSPORT ION PD000141:F319-Y486	BLAST_PRODOM
					POTASSIUM CHANNEL SIGNATURE PR00169:F319-V339, M363-C389, E392-E415, F427-M449, G456-F482, E211-P230, P245-T273, I293-K316	BLIMPS_PRINTS
					transmembrane domain: I253-C270, V356-A373, V394-L413	HMMER
					Ion transport protein ion_trans:I263-I478	HMMER_PFAM

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
18	7480826CD1	506	S12 S22 S280 S320 T125 T181 T276 T349 T433	N254 N258 N27 N274 N278 N326 N79	TRANSPORTER PROTEIN PD138374:H360-H506	BLAST_PRODOM
					ACID AMINO PROTEIN TRANSPORTER PERMEASE TRANSMEMBRANE INTERGENIC REGION PUTATIVE PROLINE PD001875:S76-I394	BLAST_PRODOM
					transmembrane domain: A97-L116, L224-V243, L192-S210, I330-T349, V375-F392, I416-I441, I473-I493	HMMER
					Transmembrane amino acid transporter protein Aa_trans:A95-S489	HMMER_PFAM

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
19	6025572CD1	315	S53 T209 T245		MITOCHONDRIAL ENERGY TRANSFER PROTEINS DM00026|S31935|110-208:Q120-K218	BLAST_DOMO
					MITOCHONDRIAL ENERGY TRANSFER PROTEINS DM00026|P02722|11-96:L25-L110	BLAST_DOMO
					PROTEIN TRANSPORT TRANSMEMBRANE REPEAT MITOCHONDRION CARRIER MEMBRANE INNER MITOCHONDRIAL ADP/ATP PD000117:S18-V210	BLAST_PRODOM
					Mitochondrial energy transfer proteins BL00215:L25-Q49, I271-G283	BLIMPS_BLOCKS
					MITOCHONDRIAL CARRIER PROTEINS PR00926:A229-M251, D23-T36, T36-V50, G85-D105, T138-D156, Y186-F204	BLIMPS_PRINTS
					ADENINE NUCLEOTIDE TRANSLOCATOR PR00927:F20-A32, Y63-R84, T96-K108, R123-G136, R164-L185, S225-Y241, E275-R290	BLIMPS_PRINTS
					Mitochondrial carrier proteins mito_carr:S19-F308	HMMER_PFAM
					Mitoch_Carrier: P40-L48, P145-L153, P242-M250	MOTIFS
					Mitochondrial energy transfer proteins signature mitoch_carrier.prf:F20-I73, F125-I176, F222-I271	PROFILESCAN

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
20	5686561CD1	540	S162 S180 S24 S29 S327 S349 S454 T527	N399 N406	Transmembrane domains: A77-Y100, Y220-L243, I259-L285, V291-Y311, A369-F389	HMMER
					Sodium channel signature: PR00170:G362-F389, Y76-G105, L361-F389, K109-G134	BLIMPS-PRINTS
					Calcium channel: DM00043|A55645|1137-1259: A250-V298 (P-value = 2.7e-5)	BLAST-DOMO
					Voltage gated calcium channel PD000032:Y221-G391, I460-F486, N423-W443 (P-value = 1.1e-6)	BLAST-PRODOM
21	1553725CD1	322	S142 S217 S295 S39 T133 T168 T304 T62 Y315	N123 N131 N29	PROTEIN TRANSMEMBRANE CHROMOSOME PUTATIVE TRANSPORTER C17G6.15C TRANSPORT XV READING FRAME PD006986:F8-L253	BLAST_PRODOM

Table 3 (cont.)

<table>
<tr><th>SEQ ID NO:</th><th>Incyte Polypeptide ID</th><th>Amino Acid Residues</th><th>Potential Phosphorylation Sites</th><th>Potential Glycosylation Sites</th><th>Signature Sequences, Domains and Motifs</th><th>Analytical Methods and Databases</th></tr>
<tr><td>22</td><td>1695770CD1</td><td>417</td><td>S108 S122 S163 S43 S56 T196 T239 T243 T410 T411 T88</td><td>N72</td><td>Signal peptide: M1-A28</td><td>HMMER</td></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td>Transmembrane domains: M255-I279, I320-I339</td><td>HMMER</td></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td>Neurotransmitter-gated ion-channel domain: P44-F341</td><td>HMMER_PFAM</td></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td>Neurotransmitter-gated ion channels signature BL00236: V73-R110, I127-N136, N157-Y195, F242-A283</td><td>BLIMPS_BLOCKS</td></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td>Neurotransmitter-gated ion-channels signature: L152-E206</td><td>PROFILESCAN</td></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td>Neurotransmitter-gated ion-channel family signature PR00252:R93-Y109, S126-E137, C172-C186, F249-Q261</td><td>BLIMPS_PRINTS</td></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td>Gamma-aminobutyric acid A (GABAA) receptor signature PR00253:Y258-W278, A284-S305, I318-I339</td><td>BLIMPS_PRINTS</td></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td>CHANNEL IONIC TRANSMEMBRANE GLYCOPROTEIN POSTSYNAPTIC MEMBRANE RECEPTOR PRECURSOR SIGNAL PROTEIN PD000153: R99-K347</td><td>BLAST_PRODOM</td></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td>NEUROTRANSMITTER-GATED ION-CHANNELS DM00560|S18836|18-453: R24-D417</td><td>BLAST_DOMO</td></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td>Neurotransmitter-gated ion channel motif: C172-C186</td><td>MOTIFS</td></tr>
</table>

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
23	4672222CD1	1864	S103 S195 S196 S2 S22 S406 S5 S547 S697 S727 S757 S836 S87 S883 T115 T12 T299 T318 T349 T367 T508 T523 T529 T593 T603 T615 T675 T778 T795 T842 Y327 S1476 S1503 T1163 S1191 S1361 S1413 T1430 S1493 S1526 S1555 S1614 T1631 S1633 T1742 T1758 S1850 T1245 S1410 S1456 T1471 S1499 S1698 S1859 Y1220 Y1552	N404 N550 N715 N718 N805 N925 N1058 N1465 N1466 N1595 N1773 N1849	Transmembrane domains: F858-M878, N999-L1022, V1079-Q1102	HMMER
					PROTEIN MELASTATIN CHROMOSOME TRANSMEMBRANE C05C12.3 T01H8.5 I F54D1.5 IV PD018035:Y108-L439 PD039592:E597-N801 PD151509:V974-P1063, W1030-K1253 PD022180:W434-R545	BLAST_PRODOM

Table 3 (cont.)

<table>
<tr><th>SEQ ID NO:</th><th>Incyte Polypeptide ID</th><th>Amino Acid Residues</th><th>Potential Phosphorylation Sites</th><th>Potential Glycosylation Sites</th><th>Signature Sequences, Domains and Motifs</th><th>Analytical Methods and Databases</th></tr>
<tr><td>24</td><td>6176128CD1</td><td>1237</td><td>S102 S135 S139 S168 S179 S361 S407 S438 S439 S538 S686 S690 S713 S720 S726 S770 S808 S871 S9 S924 S93 S954 T156 T302 T351 T391 T446 T517 T609 T718 T77 T994 S1090 S1098 S1219 S1013 S1030 T1146 T1155 T1190 T1231 S1125 S1215 S1221</td><td>N100 N133 N137 N279 N343 N584 N607 N682 N933 N1153</td><td>Transmembrane domains:
M155-Y177, M248-F264, L310-L330</td><td>HMMER</td></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td>CHANNEL POTASSIUM IONIC CALCIUMACTIVATED ALPHA CALCIUM SUBUNIT ACTIVATED PROTEIN LARGE
PD003090:R337-F629, I784-M889, L926-P983,
Y1003-E1033, Q1176-S1215</td><td>BLAST_PRODOM</td></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td>do CHANNEL; POTASSIUM; MSLO; ACTIVATED;
DM05442|A48206|351-1123: R337-F618,
P944-P983, Q1176-S1226</td><td>BLAST_DOMO</td></tr>
<tr><td>25</td><td>7473418CD1</td><td>539</td><td>S299 S321 T535</td><td>N533</td><td>Transmembrane domains:
V15-C38, C50-F67, F264-F282, A323-R341</td><td>HMMER</td></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td>Sodium:sulfate symporter signature:
BL01271:S451-I505, T132-I151,
M216-V240, P378-G399</td><td>BLIMPS_BLOCKS</td></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td>PROTEIN TRANSMEMBRANE TRANSPORT MEMBRANE INNER TRANSPORTER SODIUM SYMPORT OF COTRANSPORTER
PD000549:V15-V173, M216-W518</td><td>BLAST_PRODOM</td></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td>do RENAL; BOUND; PRO-SER-ALA; NA;
DM02914|S43561|28-507:R37-M159,
P199-W349, L367-T517</td><td>BLAST_DOMO</td></tr>
</table>

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
26	7474129CD1	755	S339 S353 S367 S463 S53 S572 S589 S653 S732 T128 T132 T255 T270 T277 T300 T343 T358 T362 T37 T376 T441 T664 Y225 Y347 Y587	N417 N648 N735	Transmembrane domains: V490-F507, L556-L573, P616-M642	HMMER
					Ank repeat: E179-K211, F226-S259, D305-K333	HMMER_PFAM
					VANILLOID RECEPTOR SUBTYPE 1 PD101189: Q52-L291	BLAST_PRODOM
					PROTEIN OLFACTORY CHANNEL B0212.5 T09A12.3 T10B10.7 VANILLOID RECEPTOR SUBTYPE F28H7.10 PD011151:N303-E430	BLAST_PRODOM
27	7481414CD1	301	S143 S203 S290 T136 T32		Transmembrane domain: L212-V230	HMMER
					Mitochondrial carrier proteins domain: Q8-M294	HMMER_PFAM
					Mitochondrial energy transfer proteins signature: BL00215:L214-Q238, V256-G268	BLIMPS_BLOCKS
					Mitochondrial energy transfer proteins signature: A10-G59, L107-I160, K204-A276, K213-N259	PROFILESCAN
					PROTEIN TRANSPORT TRANSMEMBRANE REPEAT MITOCHONDRION CARRIER MEMBRANE INNER MITOCHONDRIAL ADP/ATP PD000117: Y44-S241	BLAST_PRODOM
					Mitochondrial carrier protein motifs: P126-L134 P229-I237	MOTIFS

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
28	7481461CD1	515	S10 S104 S163 S257 S272 S277 S4 S474 S511 S97 T233 T250 T484	N81	Transmembrane domains: V117-F135, Y169-L191, I190-I215, G229-F245, I376-F395	HMMER
					Monocarboxylate transporter domain: A77-A455	HMMER_PFAM
					XLINKED PESTCONTAINING TRANSPORTER SOLUTE CARRIER FAMILY MONOCARBOXYLIC ACID TRANSPORTERS MEMBER PD030892:P33-V111	BLAST_PRODOM
					do PEST; TRANSPORTER; LINKED; DM05037|P36021|155-612:E63-M489	BLAST_DOMO

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
29	7472541CD1	1519	S223 S307 S432 S456 S472 S486 S498 S510 S538 S579 S628 S63 S648 S668 S701 S728 S732 S741 S756 S779 S826 S832 S903 S912 S986 T275 T341 T437 T449 T466 T495 T563 T597 T664 T674 T716 T73 T755 T805 T880 T945 T961 S1509 S1110 S1131 T1198 S1256 S1278 T1431 S1480 S1406 T1439 S1505 Y1079	N148 N298 N339 N354 N41 N51 N69 N991 N1249 N1331	Transmembrane domains: M313-G331, L358-L383, L1317-C1337	HMMER
					E1-E2 ATPase domains: E422-V444, L935-H985	HMMER_PFAM
					E1-E2 ATPases phosphorylation site BL00154:G173-L190, I427-F445, D949-L989	BLIMPS_BLOCKS
					E1-E2 ATPases phosphorylation site: I413-A461	PROFILESCAN
					P-type cation-transporting ATPase PR00119:F431-F445, A965-D975, I1111-I1130	BLIMPS_PRINTS
					ATPASE HYDROLASE TRANSMEMBRANE PHOSPHORYLATION ATPBINDING PROTEIN PROBABLE CALCIUMTRANSPORTING CALCIUM TRANSPORT PD004657:A1145-F1374 PD006317:Y162-E255 PD149930:C1085-F1144 PD004932:R65-P121	BLAST_PRODOM
					do ATPASE; CALCIUM; TRANSPORTING; DM02405|P32660|318-1225:R157-E475, E776-N1209	BLAST_DOMO
					E1-E2 ATPase motif: D433-T439	MOTIFS
					ATP/GTP binding site (P-loop): G1053-T1060	MOTIFS

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
30	6999183CD1	1585	S2, S7, T30, S61, S70, T86, S114, S198, T267, T459, T464, T576, S626, T717, T744, S756, T757, S779, S789, T793, S885, S893, Y917 S924, T938, T967, T971, S1005, S1054, S1097, S1158, S1202, S1262, T1267, T1296, T1339, T1381, T1410, T1427, T1431, S1457, Y1544 S1574, S1549,	N72, N121, N196, N245, N457, N546, N557, N881, N910, N960, N1272, N1337	ABC TRANSPORTERS FAMILY: DM00008|P41233|839-1045: I1268-M1455, I482-P611, E598-N652	BLAST-DOMO
					ABC transporters family: BL00211: L516-L527, L1382-D1413	BLIMPS-BLOCKS
					Transmembrane domain (transmem_domain): I1058-L1082, I1099-L1117, G1124-I1147, L1167-M1193, T30-F48, T224-V242, W271-I289, T306-I326, P329-L346, F358-M375, Y398-Y420, V1034-F1053	HMMER
					ABC transporter (ABC_tran): G511-G653, G1280-G1458	HMMER-PFAM
					ATP/GTP-binding site motif A (P-loop) (Atp_Gtp_A): G518-T525, G1287-S1294	MOTIFS
					ABC transporters family signature (atp_bind_transport.prf): I1362-D1413	PROFILESCAN

Table 4

Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID	Sequence Length	Selected Fragment(s)	Sequence Fragments	5' Position	3' Position
31	2194064CB1	1129	1071-1129, 833-898	g5110579	1	485
				FL2194064_g7770598_000019_g7670446	203	1129
				6542780F9 (LNODNON02)	32	481
32	2744094CB1	2699	1-2196, 2541-2587	FL097646_00001	431	2542
				55058921H1	1	793
				70317743D1	2347	2699
				70317681D1	2209	2639
33	2798241CB1	6369	1-1210, 1759-5012	71911330V1	5832	6369
				70300809D1	5128	5690
				6340750H1 (BRANDIN01)	5650	6322
				7601441J1 (ESOGTME01)	4623	5186
				6314138H1 (NERDTDN03)	5235	5750
				7690596H1 (PROSTME06)	4145	4636
				7753104J1 (HEAONOE01)	5764	6357
				4013186F9 (MUSCNOT10)	3758	4391
				7606344H1 (COLRTUE01)	1764	2219
				6913644H1 (PITUDIR01)	4608	5181
				55052455J1	1981	2827
				7400061H1 (SINIDME01)	1	502
				2798241T6 (NPOLNOT01)	1325	1955
				55058989J1	2548	3298
				7100413F7 (BRAWTDR02)	483	1185
				6744456H1 (BRAFNOT02)	568	1274
				55053647J1	3011	3823
				6586921H1 (TLYMUNT03)	1157	1724
34	3105257CB1	2558	1-587, 2435-2558	70864718V1	1864	2353
				70549000V1	1608	2310
				FL3105257CT1_00001	1	1843
				6451207H1 (BRAINOC01)	1868	2558

Table 4 (cont.)

Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID	Sequence Length	Selected Fragment(s)	Sequence Fragments	5' Position	3' Position
35	3200979CB1	5065	5030-5065, 1-3313	FL3200979_g3810670_g4240130	1	4779
				71698878V1	4463	5065
36	6754139CB1	1677	1-686	656293H1 (EOSINOT03)	532	800
				55062573H1	789	875
				GBI:edit	1	531
				GNN:g8017750_000028_004	386	1677
				g5678193	684	883
				6754139J1 (SINTFER02)	684	874
37	6996659CB1	3714	1-1916, 3071-3091, 2092-2619	6996659F8 (BRAXTDR17)	1180	1915
				GBI.g9211864_01_04_05_12.edit	1303	3006
				55098348H2	2752	2942
				1596150T6 (BRAINOT14)	3116	3707
				7124651F6 (COLNDIY01)	2605	2776
				g4622477	3322	3714
				1596150F6 (BRAINOT14)	2967	3466
				55063531J1	1	309
				7291716R6 (BRAIFER06)	510	1209
				7291716F6 (BRAIFER06)	219	1174
				55063924J1	1768	1994
38	7472747CB1	1009	1-388, 571-704, 778-1009	FL7472747_g6983242_000026_g3925427	122	1009
				7616162H1 (COLNTUN03)	1	450
39	7474121CB1	1155	1-1155	GNN.g7259672_000014_002	1	1155
40	7475615CB1	2733	1852-2185, 1484-1579, 665-1340, 1-249, 2334-2733, 454-495	FL7475615_g8980204_000002_g2654005_1_11-12	1580	1756
				FL7475615_g8980204_000002_g2654005_1_6-7	986	1221
				FL7475615_g8980204_000002_g2654005_1_12-13	1687	1849

Table 4 (cont.)

Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID	Sequence Length	Selected Fragment(s)	Sequence Fragments	5' Position	3' Position
				FL7475615_g8980204_000002_g2654005_1_7-8	1139	1369
				55092029J1	341	1088
				55083049H1	1	470
				1509180F6 (LUNGNOT14)	1744	2228
				FL7475615_g8980204_000002_g2654005_1_8-9	1222	1483
				GNN.g7342135_000012_002	821	1579
				6806177J1 (SKIRNOR01)	1995	2733
				FL7475615_g8980204_000002_g2654005_1_10-11	1484	1686
41	7475656CB1	3457	3284-3346, 1169-1646, 1-290, 2835-2868, 2018-2292, 3030-3174, 780-844, 456-653	55073909H2	1	110
				g3168873_CD	382	2628
				7946572H1 (BRABNOE02)	228	536
				5373417T9 (BRAINOT22)	2245	2901
				GNN.g6532090_000006_000019.edit	43	867
				2428507R6 (SCORNON02)	2911	3457
				5373417F8 (BRAINOT22)	1396	1620
				5893974H1 (BRAYDIN03)	2879	3160
				4787380H1 (BRATNOT03)	1461	1717
42	7480632CB1	5622	1-3676, 5557-5622	1450339F1 (PENITUT01)	4134	4646
				7270152H1 (OVARDIJ01)	167	646
				71697049V1	4737	5460
				3488927H1 (EPIGNOT01)	2908	3106
				6774619J1 (OVARDIR01)	2052	2720
				5063703F6 (ARTFTDT01)	4886	5622
				55072886J1	3675	4262
				GBI.g3810670_000001.edit	266	4674
				6488228F9 (MIXDUNB01)	1	632
				7670233H2 (BONRNOC01)	4317	4878

Table 4 (cont.)

Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID	Sequence Length	Selected Fragment(s)	Sequence Fragments	5' Position	3' Position
43	6952742CB1	2600	2329-2600, 1-224, 1190-1560, 1957-2046, 1006-1030	5884027F8 (LIVRNON08)	2088	2600
				55063579H1	779	1276
				GBI:g7232144_000013.edit.3	1626	2351
				6816048H1 (ADRETUR01)	1	316
				6952742H1 (BRAITDR02)	1140	1824
				GNN.g6970605_000013_002	342	1355
				GBI:g7232144_000013.fasta.edit	255	506
44	7478795CB1	2917	2698-2917, 1808-2065, 398-714, 923-976	72016954V1	2193	2917
				71989431V1	2165	2912
				72017820V1	1369	2155
				72017055V1	1160	2053
				72017371V1	570	1212
				72017076V1	1958	2859
				72017430V1	476	1146
				55076285H1	1	566
45	656293CB1	1474	1-362	GBI.g8017750_edit	1	1353
				FL656293_g8017750_000028_g6746563_2_2-3	130	895
				FL656293_g8017750_000028_g6746563_2_3-4	363	1353
				7675576H1 (NOSETUE01)	907	1474
46	7473957CB1	1742	1-367, 1680-1742	4648731F9 (PROSTUT20)	610	1274
				71166638V1	1	610
				71165785V1	1010	1742
				6830443J1 (SINTNOR01)	592	1266
47	7474111CB1	2312	1-639, 1686-1712, 2004-2312, 1860-1908	7761487H1 (THYMNOE02)	6	632
				6770140H1 (BRAUNOR01)	1692	2312
				7761487J1 (THYMNOE02)	1	506
				GNN.g7243948_CDS_1	183	1845

Table 4 (cont.)

Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID	Sequence Length	Selected Fragment(s)	Sequence Fragments	5' Position	3' Position
48	7480826CB1	2320	161-224, 2044-2320	7752763J1 (HEAONOE01)	1668	2320
				60143671D1	467	917
				6052064J1 (BRABDIR03)	1080	1658
				6484950H1 (MIXDUNB01)	1276	1723
				2944045H1 (BRAITUT23)	827	1118
				7469461H1 (LUNGNOE02)	1	498
49	6025572CB1	1781	1-170	FL6025572_g7382154_000015_g1 197164	347	1063
				4923834H1 (TESTNOT11)	1	291
				g3838735	1313	1781
				g3734777	252	472
				71970611V1	1285	1780
				6025572F6 (TESTNOT11)	883	1627
50	5686561CB1	2433	1-1078, 1197-1275	71412362V1	1088	1702
				6060785H1 (BRAENOT04)	551	1100
				7695065J1 (LNODTUE01)	387	1052
				7633409H1 (SINTDIE01)	1	483
				3776733H1 (BRSTNOT27)	2148	2433
				2802364F6 (PENCNOT01)	1765	2304
				5564984F6 (TLYMNOT08)	860	1528
				70730430V1	1525	2108
51	1553725CB1	1772	1571-1772	60211064U1	344	823
				72050509V1	1176	1772
				70300327D1	984	1428
				70300706D1	1	262
				1553725X15C1 (BLADTUT04)	54	694
				70300332D1	729	1286
52	1695770CB1	1874	1-479, 1298-1874, 1131-1216, 886-984	55117454H1	1155	1874
				55110123H1	286	1179
				55072985J1	1	542

Table 4 (cont.)

Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID	Sequence Length	Selected Fragment(s)	Sequence Fragments	5' Position	3' Position
53	4672222CB1	6211	3238-3683, 4625-4798, 2313-2462, 1-1636	55047368J1	1925	2815
				71007436V1	5663	6211
				71998604V1	4613	5344
				71995592V1	3913	4598
				3462433F7 (293TF2T01)	2738	3162
				71997753V1	4522	5239
				71995863V1	3346	3886
				55073038H1)	818	1499
				55141177J1	2942	3318
				71998657V1	3811	4537
				6141577F6 (BMARTXT03)	1	878
				55140386J1	1086	1915
				GBI:g8189326.edit	2957	3903
				5092011F6 (UTRSTMR01)	1797	2436
				7743692H1 (ADRETUE04)	5374	5927
				2505959F6 (CONUTUT01)	5325	5866
54	6176128CB1	3714	1-197, 329-2513, 3301-3336	GBI.g979669_000005_000004.edit	1	1143
				6859776H1 (BRAIFEN08)	2265	2953
				GBI.g979669_000002.edit	3612	3714
				GBI.g7739135_000005.edit	3115	3711
				6772216J1 (BRAUNOR01)	2991	3324
				6887873J1 (BRAITDR03)	899	1503
				8039114H1 (SPLNNOE01)	1741	2374
				6907605J1 (PITUDIR01)	2586	3088
				6445788H2 (BRAINOC01)	1383	2006
				6891702F6 (BRAITDR03)	543	1053
				7065904R6 (BRATNOR01)	383	645
55	7473418CB1	3115	1-1411	FL7473418_g3176728_g5531902_1_4-5	369	740
				7056016H1 (BRALNON02)	2658	3115

Table 4 (cont.)

Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID	Sequence Length	Selected Fragment(s)	Sequence Fragments	5' Position	3' Position
				FL7473418_g3176728_g5531902_1_7-8	864	1188
				1324158F6 (LPARNOT02)	2114	2695
				6899347H1 (LIVRTMR01)	1074	1568
				FL7473418_g3176728_g5531902_1_5-6	548	863
				70075691U1	2291	2782
				FL7473418_g3176728_g5531902_1_6-7	739	1069
				FL7473418_g3176728_g5531902_1_1-2	1	231
				FL7473418_g3176728_g5531902_1_10-11	1351	1620
				FL7473418_g3176728_g5531902_1_2-3	103	368
				7114876H1 (BRAENOK01)	1549	1954
				FL7473418_g3176728_g5531902_1_3-4	232	547
				4895008F6 (LIVRTUT12)	1752	2240
56	7474129CB1	2846	1-1696, 2073-2846, 1777-2012	55109928H1	2480	2846
				55109306J1	1837	2660
				55124533H1	1	832
				55124525H1	1073	1893
				55073088J1	796	1208
57	7481414CB1	906	441-541, 262-348	GBI.g9454493_000005_000056.edit	1	906

Table 4 (cont.)

Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID	Sequence Length	Selected Fragment(s)	Sequence Fragments	5' Position	3' Position
58	7481461CB1	1840	1-91	70481006V1	551	1137
				70465445V1	673	1250
				1748722F6 (STOMTUT02)	1423	1840
				60266587D1	1225	1801
				7637372J1 (SINTDIE01)	221	642
				g5933739	1	407
59	7472541CB1	5348	1384-1560, 1-1188, 4239-4906, 2145-2970, 4944-5348	6772907J1 (BRAUNOR01)	1641	2204
				2182261F6 (SININOT01)	4301	4634
				5459667H1 (SINITUT03)	3276	3550
				GNN.g7710567_000006_002.edit	1057	2001
				7362215H1 (BRAIFEE05)	1	526
				GNN.g7708823_000019_002	4306	4862
				7032970H1 (BRAXTDR12)	2692	3426
				7313608H1 (BRABDIE02)	731	1260
				71462931V1	4732	5348
				5767060H1 (STOMFET02)	3887	4428
				8069315J1 (BRAIFEE05)	91	792
				7582660H1 (BRAIFEC01)	3510	4131
				GNN.g7454125_000004_002.edit	1382	3918
				6772907H1 (BRAUNOR01)	2476	2994
60	6999183CB1	5149	1-1797, 4753-4852, 3028-3711, 2471-2667	GBI.g3873182_000001.edit5p	1	3167
				72017145V1	4087	4903
				6999183R8 (HEALDIR01)	384	1128
				55051672H1	909	1573
				72017349V1	3836	4767
				72293922V1	2616	3434
				55144835H1	2154	2939
				55144834J1	3263	3947
				55076606J1	1126	1592
				72017610V1	4219	5149

Table 5

Polynucleotide SEQ ID NO:	Incyte Project ID	Representative Library
31	2194064CB1	THYRTUT03
32	2744094CB1	BRSTTUT15
33	2798241CB1	PROSTME06
34	3105257CB1	BLADNOT01
35	3200979CB1	PENITUT01
36	6754139CB1	BRSTNOR01
37	6996659CB1	BRAIFER06
38	7472747CB1	COLNTUN03
40	7475615CB1	LUNGNON07
41	7475656CB1	BRAINOT22
42	7480632CB1	PENITUT01
43	6952742CB1	LIVRNON08
44	7478795CB1	BRAENOT02
45	656293CB1	COLNNOT22
46	7473957CB1	BRAHTDR03
47	7474111CB1	THYMNOE02
48	7480826CB1	MIXDUNB01
49	6025572CB1	TESTNOT11
50	5686561CB1	BRAENOT04
51	1553725CB1	THYMNON04
52	1695770CB1	COLNNOT23
53	4672222CB1	PITUDIR01
54	6176128CB1	BRAITDR03
55	7473418CB1	LPARNOT02
56	7474129CB1	PLACNOT05
58	7481461CB1	OVARTUT05
59	7472541CB1	BRAIFEE05
60	6999183CB1	HEALDIR01

Table 6

Library	Vector	Library Description
BLADNOT01	PBLUESCRIPT	Library was constructed using RNA isolated from the bladder tissue of a 78-year-old Caucasian female, who died from an intracranial bleed. Patient history included basal cell carcinoma, arthritis, and chronic hypertension.
BRAENOT02	pINCY	Library was constructed using RNA isolated from posterior parietal cortex tissue removed from the brain of a 35-year-old Caucasian male who died from cardiac failure.
BRAENOT04	pINCY	Library was constructed using RNA isolated from inferior parietal cortex tissue removed from the brain of a 35-year-old Caucasian male who died from cardiac failure. Pathology indicated moderate leptomeningeal fibrosis and multiple microinfarctions of the cerebral neocortex. Patient history included dilated cardiomyopathy, congestive heart failure, cardiomegaly and an enlarged spleen and liver .
BRAHTDR03	PCDNA2.1	This random primed library was constructed using RNA isolated from archaecortex, anterior hippocampus tissue removed from a 55-year-old Caucasian female who died from cholangiocarcinoma. Pathology indicated mild meningeal fibrosis predominately over the convexities, scattered axonal spheroids in the white matter of the cingulate cortex and the thalamus, and a few scattered neurofibrillary tangles in the entorhinal cortex and the periaqueductal gray region. Pathology for the associated tumor tissue indicated well-differentiated cholangiocarcinoma of the liver with residual or relapsed tumor. Patient history included cholangiocarcinoma, post-operative Budd-Chiari syndrome, biliary ascites, hydorthorax, dehydration, malnutrition, oliguria and acute renal failure. Previous surgeries included cholecystectomy and resection of 85% of the liver.
BRAIFEE05	PCDNA2.1	This 5' biased random primed library was constructed using RNA isolated from brain tissue removed from a Caucasian male fetus who was stillborn with a hypoplastic left heart at 23 weeks' gestation.
BRAIFER06	PCDNA2.1	This random primed library was constructed using RNA isolated from brain tissue removed from a Caucasian male fetus who was stillborn with a hypoplastic left heart at 23 weeks' gestation. Serologies were negative.

Table 6 (cont.)

Library	Vector	Library Description
BRAINOT22	pINCY	Library was constructed using RNA isolated from right temporal lobe tissue removed from a 45-year-old Black male during a brain lobectomy. Pathology for the associated tumor tissue indicated dysembryoplastic neuroepithelial tumor of the right temporal lobe. The right temporal region dura was consistent with calcifying pseudotumor of the neuraxis. Family history included obesity, benign hypertension, cirrhosis of the liver, obesity, hyperlipidemia, cerebrovascular disease, and type II diabetes.
BRAITDR03	PCDNA2.1	This random primed library was constructed using RNA isolated from allocortex, cingulate posterior tissue removed from a 55-year-old Caucasian female who died from cholangiocarcinoma. Pathology indicated mild meningeal fibrosis predominately over the convexities, scattered axonal spheroids in the white matter of the cingulate cortex and the thalamus, and a few scattered neurofibrillary tangles in the entorhinal cortex and the periaqueductal gray region. Pathology for the associated tumor tissue indicated well-differentiated cholangiocarcinoma of the liver with residual or relapsed tumor. Patient history included cholangiocarcinoma, post-operative Budd-Chiari syndrome, biliary ascites, hydorthorax, dehydration, malnutrition, oliguria and acute renal failure. Previous surgeries included cholecystectomy and resection of 85% of the liver.
BRSTNOR01	pINCY	Library was constructed using RNA isolated from diseased breast tissue removed from a 59-year-old Caucasian female during a unilateral extended simple mastectomy. Pathology for the associated tumor tissue indicated an invasive lobular carcinoma with extension into ducts. Patient history included cirrhosis, esophageal ulcer, hyperlipidemia, and neuropathy.
BRSTTUT15	pINCY	Library was constructed using RNA isolated from breast tumor tissue removed from a 46-year-old Caucasian female during a unilateral extended simple mastectomy.. Pathology indicated invasive grade 3, nuclear grade 2 adenocarcinoma, ductal type. An intraductal carcinoma component, non- comedo, comprised approximately 50% of the neoplasm, including the lactiferous ducts. Angiolymphatic involvement was present. Metastatic adenocarcinoma was present in 7 of 10 axillary lymph nodes. The largest nodal metastasis measured 3 cm, and focal extracapsular extension was identified. Family history included atherosclerotic coronary artery disease, type II diabetes, cerebrovascular disease, and depressive disorder.

Table 6 (cont.)

Library	Vector	Library Description
COLNNOT22	pINCY	Library was constructed using RNA isolated from colon tissue removed from a 56-year-old Caucasian female with Crohn's disease during a partial resection of the small intestine. Pathology indicated Crohn's disease of the ileum and ileal-colonic anastomosis, causing a fistula at the anastomotic site that extended into pericolonic fat. The ileal mucosa showed linear and puncture ulcers with intervening normal tissue. Previous surgeries included a partial ileal resection and permanent ileostomy. Family history included irritable bowel syndrome in the mother and the siblings.
COLNNOT23	pINCY	Library was constructed using RNA isolated from diseased colon tissue removed from a 16-year-old Caucasian male during a total colectomy with abdominal/perineal resection. Pathology indicated gastritis and pancolonitis consistent with the acute phase of ulcerative colitis. Inflammation was more severe in the transverse colon, with inflammation confined to the mucosa. There was only mild involvement of the ascending and sigmoid colon, and no significant involvement of the cecum, rectum, or terminal ileum. Family history included irritable bowel syndrome.
COLNTUN03	pINCY	This normalized pooled colon tumor tissue library was constructed from 1.16 million independent clones from a pooled colon tumor library. Starting library was constructed using pooled cDNA from 6 donors. cDNA was generated using mRNA isolated from colon tumor tissue removed from a 55-year-old Caucasian male (A) during hemicolectomy; from a 60-year-old Caucasian male (B) during hemicolectomy; from a 62-year-old Caucasian male (C) during sigmoidectomy; from a 30-year-old Caucasian female (D) during hemicolectomy; from a 64-year-old Caucasian female (E) during hemicolectomy; and from a 70-year-old Caucasian female (F) during hemicolectomy. Pathology indicated invasive grade 3 adenocarcinoma (A); invasive grade 2 adenocarcinoma (B); invasive grade 2 adenocarcinoma (C); carcinoid tumor (D); invasive grade 3 adenocarcinoma (E); and invasive grade 2 adenocarcinoma (F). Donors B, C, D, E, and F had positive lymph nodes. Patient medications included Ativan (A); Seldane (B), Tri-Levlen (D); Synthroid (E); Tamoxifen, prednisone, Synthroid, and Glipizide (F). The library was normalized in two rounds using conditions adapted from Soares et al., PNAS (1994) 91:9228-9232 and Bonaldo et al., Genome Research 6 (1996):791, except that a significantly longer (48 hours/round) reannealing hybridization was used.

Table 6 (cont.)

Library	Vector	Library Description
HEALDIR01	PCDNA2.1	This random primed library was constructed using RNA isolated from diseased left ventricle tissue removed from a 7-month-old Caucasian male who died from cardiopulmonary arrest due to Pompe's disease. Patient history included Pompe's disease, left ventricular hypertrophy, pyrexia, right complete cleft lip, cleft palate, chronic serous otitis media, hypertrophic cardiomyopathy, congestive heart failure, and developmental delays. Family history included acute myocardial infarction, diabetes, cystic fibrosis and Down's syndrome.
LIVRNON08	pINCY	This normalized library was constructed from 5.7 million independent clones from a pooled liver tissue library. Starting RNA was made from pooled liver tissue removed from a 4-year-old Hispanic male who died from anoxia and a 16 week female fetus who died after 16-weeks gestation from anencephaly. Serologies were positive for cytolomegalovirus in the 4-year-old. Patient history included asthma in the 4-year-old. Family history included taking daily prenatal vitamins and mitral valve prolapse in the mother of the fetus. The library was normalized in 2 rounds using conditions adapted from Soares et al., PNAS (1994) 91:9228 and Bonaldo et al., Genome Research 6 (1996):791, except that a significantly longer (48 hours/round) reannealing hybridization was used.
LPARNOT02	pINCY	Library was constructed using RNA isolated from tissue obtained from the left parotid (salivary) gland of a 70-year-old male with parotid cancer.
LUNGNON07	pINCY	This normalized lung tissue library was constructed from 5.1 million independent clones from a lung tissue library. Starting RNA was made from RNA isolated from lung tissue. The library was normalized in two rounds using conditions adapted from Soares et al., PNAS (1994) 91:9228-9232 and Bonaldo et al., Genome Research (1996) 6:791, except that a significantly longer (48 hours/round) reannealing hybridization was used.

Table 6 (cont.)

Library	Vector	Library Description
MIXDUNB01	pINCY	Library was constructed using RNA isolated from myometrium removed from a 41-year-old Caucasian female (A) during vaginal hysterectomy with a dilatation and curettage and untreated smooth muscle cells removed from the renal vein of a 57-year-old Caucasian male. Pathology for donor A indicated the myometrium and cervix were unremarkable. The endometrium was secretory and contained fragments of endometrial polyps. Benign endo- and ectocervical mucosa were identified in the endocervix. Pathology for the associated tumor tissue indicated uterine leiomyoma. Medical history included an unspecified menstrual disorder, ventral hernia, normal delivery, a benign ovarian neoplasm, and tobacco abuse in donor A. Previous surgeries included a bilateral destruction of fallopian tubes, removal of a solitary ovary, and an exploratory laparotomy in donor A. Medications included ferrous sulfate in donor A.
OVARTUT05	pINCY	Library was constructed using RNA isolated from ovarian tumor tissue removed from a 62-year-old Caucasian female during a total abdominal hysterectomy, removal of the fallopian tubes and ovaries, exploratory laparotomy, regional lymph node excision, and dilation and curettage. Pathology indicated a grade 4 endometrioid carcinoma with extensive squamous differentiation, forming a solid mass in the right ovary. The uterine endometrium was inactive, the cervix showed mild chronic cervicitis, and focal endometriosis was observed in the posterior uterine serosa. Curettings indicated weakly proliferative endometrium with excessive stromal breakdown in the uterus, and a prior cervical biopsy indicated mild chronic cervicitis with a prominent nabothian cyst in the cervix. Patient history included longitudinal deficiency of the radioulna, osteoarthritis, thrombophlebitis, and abnormal blood chemistries. Family history included atherosclerotic coronary artery disease, pulmonary embolism, and cerebrovascular disease.
PENITUT01	pINCY	Library was constructed using RNA isolated from tumor tissue removed from the penis of a 64-year-old Caucasian male during penile amputation. Pathology indicated a fungating invasive grade 4 squamous cell carcinoma involving the inner wall of the foreskin and extending onto the glans penis. Patient history included benign neoplasm of the large bowel, atherosclerotic coronary artery disease, angina pectoris, gout, and obesity. Family history included malignant pharyngeal neoplasm, chronic lymphocytic leukemia, and chronic liver disease.

Table 6 (cont.)

Library	Vector	Library Description
PITUDIR01	PCDNA2.1	This random primed library was constructed using RNA isolated from pituitary gland tissue removed from a 70-year-old female who died from metastatic adenocarcinoma.
PLACNOT05	pINCY	Library was constructed using RNA isolated from placental tissue removed from a Caucasian male fetus, who died after 18 weeks' gestation from fetal demise.
PROSTME06	PCDNA2.1	This 5' biased random primed library was constructed using RNA isolated from diseased prostate tissue removed from a 57-year-old Caucasian male during closed prostatic biopsy, radical prostatectomy, and regional lymph node excision. Pathology indicated adenofibromatous hyperplasia. Pathology for the matched tumor tissue indicated adenocarcinoma, Gleason grade 3+3, forming a predominant mass involving the right side centrally. The patient presented with elevated prostate specific antigen and prostate cancer. Patient history included tobacco abuse in remission. Previous surgeries included cholecystectomy, repair of diaphragm hernia, and repair of vertebral fracture. Patient medications included Pepsid, Omnipen, and Eulexin. Family history included benign hypertension, cerebrovascular accident, atherosclerotic coronary artery disease, uterine cancer and type II diabetes in the mother; prostate cancer in the father; drug abuse, prostate cancer, and breast cancer in the sibling(s).
TESTNOT11	pINCY	Library was constructed using RNA isolated from testicular tissue removed from a 16-year-old Caucasian male who died from hanging. Patient history included drug use (tobacco, marijuana, and cocaine use), and medications included Lithium, Ritalin, and Paxil.

Table 6 (cont.)

Library	Vector	Library Description
THYMNOE02	PCDNA2.1	This 5' biased random primed library was constructed using RNA isolated from thymus tissue removed from a 3-year-old Hispanic male during a thymectomy and closure of a patent ductus arteriosus. The patient presented with severe pulmonary stenosis and cyanosis. Patient history included a cardiac catheterization and echocardiogram. Previous surgeries included Blalock-Taussig shunt and pulmonary valvotomy. The patient was not taking any medications. Family history included benign hypertension, osteoarthritis, depressive disorder, and extrinsic asthma in the grandparent(s).
THYMNON04	PSPORT1	This normalized library was constructed from a thymus tissue library. Starting RNA was made from thymus tissue removed from a 3-year-old Caucasian male, who died from anoxia. Serologies were negative. The patient was not taking any medications. The library was normalized in two rounds using conditions adapted from Soares et al., PNAS (1994) 91:9228 and Bonaldo et al., Genome Research (1996) 6:791, except that a significantly longer (48-hours/round) reannealing hybridization was used.
THYRTUT03	pINCY	Library was constructed using RNA isolated from benign thyroid tumor tissue removed from a 17-year-old Caucasian male during a thyroidectomy. Pathology indicated encapsulated follicular adenoma forming a circumscribed mass.

Table 7

Program	Description	Reference	Parameter Threshold
ABI FACTURA	A program that removes vector sequences and masks ambiguous bases in nucleic acid sequences.	Applied Biosystems, Foster City, CA.	
ABI/PARACEL FDF	A Fast Data Finder useful in comparing and annotating amino acid or nucleic acid sequences.	Applied Biosystems, Foster City, CA; Paracel Inc., Pasadena, CA.	Mismatch <50%
ABI AutoAssembler	A program that assembles nucleic acid sequences.	Applied Biosystems, Foster City, CA.	
BLAST	A Basic Local Alignment Search Tool useful in sequence similarity search for amino acid and nucleic acid sequences. BLAST includes five functions: blastp, blastn, blastx, tblastn, and tblastx.	Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410; Altschul, S.F. et al. (1997) Nucleic Acids Res. 25:3389-3402.	*ESTs:* Probability value= 1.0E-8 or less *Full Length sequences*: Probability value= 1.0E-10 or less
FASTA	A Pearson and Lipman algorithm that searches for similarity between a query sequence and a group of sequences of the same type. FASTA comprises as least five functions: fasta, tfasta, fastx, tfastx, and ssearch.	Pearson, W.R. and D.J. Lipman (1988) Proc. Natl. Acad Sci. USA 85:2444-2448; Pearson, W.R. (1990) Methods Enzymol. 183:63-98; and Smith, T.F. and M.S. Waterman (1981) Adv. Appl. Math. 2:482-489.	*ESTs:* fasta E value=1.06E-6 *Assembled ESTs*: fasta Identity= 95% or greater and Match length=200 bases or greater; fastx E value=1.0E-8 or less *Full Length sequences:* fastx score=100 or greater
BLIMPS	A BLocks IMProved Searcher that matches a sequence against those in BLOCKS, PRINTS, DOMO, PRODOM, and PFAM databases to search for gene families, sequence homology, and structural fingerprint regions.	Henikoff, S. and J.G. Henikoff (1991) Nucleic Acids Res. 19:6565-6572; Henikoff, J.G. and S. Henikoff (1996) Methods Enzymol. 266:88-105; and Attwood, T.K. et al. (1997) J. Chem. Inf. Comput. Sci. 37:417-424.	Probability value= 1.0E-3 or less
HMMER	An algorithm for searching a query sequence against hidden Markov model (HMM)-based databases of protein family consensus sequences, such as PFAM.	Krogh, A. et al. (1994) J. Mol. Biol. 235:1501-1531; Sonnhammer, E.L.L. et al. (1988) Nucleic Acids Res. 26:320-322; Durbin, R. et al. (1998) Our World View, in a Nutshell, Cambridge Univ. Press, pp. 1-350.	*PFAM hits:* Probability value= 1.0E-3 or less *Signal peptide hits:* Score= 0 or greater

Table 7 (cont.)

Program	Description	Reference	Parameter Threshold
ProfileScan	An algorithm that searches for structural and sequence motifs in protein sequences that match sequence patterns defined in Prosite.	Gribskov, M. et al. (1988) CABIOS 4:61-66; Gribskov, M. et al. (1989) Methods Enzymol. 183:146-159; Bairoch, A. et al. (1997) Nucleic Acids Res. 25:217-221.	Normalized quality score≥GCG-specified "HIGH" value for that particular Prosite motif. Generally, score=1.4-2.1.
Phred	A base-calling algorithm that examines automated sequencer traces with high sensitivity and probability.	Ewing, B. et al. (1998) Genome Res. 8:175-185; Ewing, B. and P. Green (1998) Genome Res. 8:186-194.	
Phrap	A Phils Revised Assembly Program including SWAT and CrossMatch, programs based on efficient implementation of the Smith-Waterman algorithm, useful in searching sequence homology and assembling DNA sequences.	Smith, T.F. and M.S. Waterman (1981) Adv. Appl. Math. 2:482-489; Smith, T.F. and M.S. Waterman (1981) J. Mol. Biol. 147:195-197; and Green, P., University of Washington, Seattle, WA.	Score= 120 or greater; Match length= 56 or greater
Consed	A graphical tool for viewing and editing Phrap assemblies.	Gordon, D. et al. (1998) Genome Res. 8:195-202.	
SPScan	A weight matrix analysis program that scans protein sequences for the presence of secretory signal peptides.	Nielson, H. et al. (1997) Protein Engineering 10:1-6; Claverie, J.M. and S. Audic (1997) CABIOS 12:431-439.	Score=3.5 or greater
TMAP	A program that uses weight matrices to delineate transmembrane segments on protein sequences and determine orientation.	Persson, B. and P. Argos (1994) J. Mol. Biol. 237:182-192; Persson, B. and P. Argos (1996) Protein Sci. 5:363-371.	
TMHMMER	A program that uses a hidden Markov model (HMM) to delineate transmembrane segments on protein sequences and determine orientation.	Sonnhammer, E.L. et al. (1998) Proc. Sixth Intl. Conf. on Intelligent Systems for Mol. Biol., Glasgow et al., eds., The Am. Assoc. for Artificial Intelligence Press, Menlo Park, CA, pp. 175-182.	
Motifs	A program that searches amino acid sequences for patterns that matched those defined in Prosite.	Bairoch, A. et al. (1997) Nucleic Acids Res. 25:217-221; Wisconsin Package Program Manual, version 9, page M51-59, Genetics Computer Group, Madison, WI.	

What is claimed is:

1. An isolated polypeptide selected from the group consisting of:

a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-30,

b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-30,

c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-30, and

d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-30.

2. An isolated polypeptide of claim 1 selected from the group consisting of SEQ ID NO:1-30.

3. An isolated polynucleotide encoding a polypeptide of claim 1.

4. An isolated polynucleotide encoding a polypeptide of claim 2.

5. An isolated polynucleotide of claim 4 selected from the group consisting of SEQ ID NO:31-60.

6. A recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide of claim 3.

7. A cell transformed with a recombinant polynucleotide of claim 6.

8. A transgenic organism comprising a recombinant polynucleotide of claim 6.

9. A method of producing a polypeptide of claim 1, the method comprising:

a) culturing a cell under conditions suitable for expression of the polypeptide, wherein said cell is transformed with a recombinant polynucleotide, and said recombinant polynucleotide comprises a promoter sequence operably linked to a polynucleotide encoding the polypeptide of claim 1, and

b) recovering the polypeptide so expressed.

10. An isolated antibody which specifically binds to a polypeptide of claim 1.

11. An isolated polynucleotide selected from the group consisting of:

a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:31-60,

b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:31-60,

c) a polynucleotide complementary to a polynucleotide of a),

d) a polynucleotide complementary to a polynucleotide of b), and

e) an RNA equivalent of a)-d).

12. An isolated polynucleotide comprising at least 60 contiguous nucleotides of a polynucleotide of claim 11.

13. A method of detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide of claim 11, the method comprising:

a) hybridizing the sample with a probe comprising at least 20 contiguous nucleotides comprising a sequence complementary to said target polynucleotide in the sample, and which probe specifically hybridizes to said target polynucleotide, under conditions whereby a hybridization complex is formed between said probe and said target polynucleotide or fragments thereof, and

b) detecting the presence or absence of said hybridization complex, and, optionally, if present, the amount thereof.

14. A method of claim 13, wherein the probe comprises at least 60 contiguous nucleotides.

15. A method of detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide of claim 11, the method comprising:

a) amplifying said target polynucleotide or fragment thereof using polymerase chain reaction amplification, and

b) detecting the presence or absence of said amplified target polynucleotide or fragment thereof, and, optionally, if present, the amount thereof.

16. A composition comprising a polypeptide of claim 1 and a pharmaceutically acceptable excipient.

17. A composition of claim 16, wherein the polypeptide has an amino acid sequence selected from the group consisting of SEQ ID NO:1-30.

18. A method for treating a disease or condition associated with decreased expression of functional TRICH, comprising administering to a patient in need of such treatment the composition of claim 16.

19. A method of screening a compound for effectiveness as an agonist of a polypeptide of claim 1, the method comprising:

a) exposing a sample comprising a polypeptide of claim 1 to a compound, and

b) detecting agonist activity in the sample.

20. A composition comprising an agonist compound identified by a method of claim 19 and a pharmaceutically acceptable excipient.

21. A method for treating a disease or condition associated with decreased expression of functional TRICH, comprising administering to a patient in need of such treatment a composition of claim 20.

22. A method of screening a compound for effectiveness as an antagonist of a polypeptide of claim 1, the method comprising:

a) exposing a sample comprising a polypeptide of claim 1 to a compound, and

b) detecting antagonist activity in the sample.

23. A composition comprising an antagonist compound identified by a method of claim 22 and a pharmaceutically acceptable excipient.

24. A method for treating a disease or condition associated with overexpression of functional TRICH, comprising administering to a patient in need of such treatment a composition of claim 23.

25. A method of screening for a compound that specifically binds to the polypeptide of claim

1, the method comprising:

a) combining the polypeptide of claim 1 with at least one test compound under suitable conditions, and

b) detecting binding of the polypeptide of claim 1 to the test compound, thereby identifying a compound that specifically binds to the polypeptide of claim 1.

26. A method of screening for a compound that modulates the activity of the polypeptide of claim 1, the method comprising:

a) combining the polypeptide of claim 1 with at least one test compound under conditions permissive for the activity of the polypeptide of claim 1,

b) assessing the activity of the polypeptide of claim 1 in the presence of the test compound, and

c) comparing the activity of the polypeptide of claim 1 in the presence of the test compound with the activity of the polypeptide of claim 1 in the absence of the test compound, wherein a change in the activity of the polypeptide of claim 1 in the presence of the test compound is indicative of a compound that modulates the activity of the polypeptide of claim 1.

27. A method of screening a compound for effectiveness in altering expression of a target polynucleotide, wherein said target polynucleotide comprises a sequence of claim 5, the method comprising:

a) exposing a sample comprising the target polynucleotide to a compound, under conditions suitable for the expression of the target polynucleotide,

b) detecting altered expression of the target polynucleotide, and

c) comparing the expression of the target polynucleotide in the presence of varying amounts of the compound and in the absence of the compound.

28. A method of assessing toxicity of a test compound, the method comprising:

a) treating a biological sample containing nucleic acids with the test compound,

b) hybridizing the nucleic acids of the treated biological sample with a probe comprising at least 20 contiguous nucleotides of a polynucleotide of claim 11 under conditions whereby a specific hybridization complex is formed between said probe and a target polynucleotide in the biological sample, said target polynucleotide comprising a polynucleotide sequence of a polynucleotide of claim 11 or fragment thereof,

c) quantifying the amount of hybridization complex, and

d) comparing the amount of hybridization complex in the treated biological sample with the

amount of hybridization complex in an untreated biological sample, wherein a difference in the amount of hybridization complex in the treated biological sample is indicative of toxicity of the test compound.

29. A diagnostic test for a condition or disease associated with the expression of TRICH in a biological sample, the method comprising:

a) combining the biological sample with an antibody of claim 10, under conditions suitable for the antibody to bind the polypeptide and form an antibody:polypeptide complex, and

b) detecting the complex, wherein the presence of the complex correlates with the presence of the polypeptide in the biological sample.

30. The antibody of claim 10, wherein the antibody is:

a) a chimeric antibody,

b) a single chain antibody,

c) a Fab fragment,

d) a $F(ab')_2$ fragment, or

e) a humanized antibody.

31. A composition comprising an antibody of claim 10 and an acceptable excipient.

32. A method of diagnosing a condition or disease associated with the expression of TRICH in a subject, comprising administering to said subject an effective amount of the composition of claim 31.

33. A composition of claim 31, wherein the antibody is labeled.

34. A method of diagnosing a condition or disease associated with the expression of TRICH in a subject, comprising administering to said subject an effective amount of the composition of claim 33.

35. A method of preparing a polyclonal antibody with the specificity of the antibody of claim 10, the method comprising:

a) immunizing an animal with a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-30, or an immunogenic fragment thereof, under conditions to

elicit an antibody response,

b) isolating antibodies from said animal, and

c) screening the isolated antibodies with the polypeptide, thereby identifying a polyclonal antibody which binds specifically to a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-30.

36. An antibody produced by a method of claim 35.

37. A composition comprising the antibody of claim 36 and a suitable carrier.

38. A method of making a monoclonal antibody with the specificity of the antibody of claim 10, the method comprising:

a) immunizing an animal with a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-30, or an immunogenic fragment thereof, under conditions to elicit an antibody response,

b) isolating antibody producing cells from the animal,

c) fusing the antibody producing cells with immortalized cells to form monoclonal antibody-producing hybridoma cells,

d) culturing the hybridoma cells, and

e) isolating from the culture monoclonal antibody which binds specifically to a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-30.

39. A monoclonal antibody produced by a method of claim 38.

40. A composition comprising the antibody of claim 39 and a suitable carrier.

41. The antibody of claim 10, wherein the antibody is produced by screening a Fab expression library.

42. The antibody of claim 10, wherein the antibody is produced by screening a recombinant immunoglobulin library.

43. A method of detecting a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-30 in a sample, the method comprising:

a) incubating the antibody of claim 10 with a sample under conditions to allow specific binding of the antibody and the polypeptide, and

b) detecting specific binding, wherein specific binding indicates the presence of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-30 in the sample.

44. A method of purifying a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-30 from a sample, the method comprising:

a) incubating the antibody of claim 10 with a sample under conditions to allow specific binding of the antibody and the polypeptide, and

b) separating the antibody from the sample and obtaining the purified polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-30.

45. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:1.

46. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:2.

47. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:3.

48. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:4.

49. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:5.

50. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:6.

51. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:7.

52. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:8.

53. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:9.

54. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:10.

55. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:11.

56. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:12.

57. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:13.

58. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:14.

59. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:15.

60. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:16.

61. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:17.

62. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:18.

63. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:19.

64. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:20.

65. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:21.

66. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:22.

67. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:23.

68. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:24.

69. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:25.

70. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:26.

71. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:27.

72. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:28.

73. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:29.

74. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:30.

75. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:31.

76. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:32.

77. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:33.

78. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:34.

79. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:35.

80. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:36.

81. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:37.

82. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:38.

83. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:39.

84. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:40.

85. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:41.

86. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:42.

87. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:43.

88. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:44.

89. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:45.

90. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:46.

91. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:47.

92. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:48.

93. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:49.

94. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:50.

95. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:51.

96. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID

NO:52.

97. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:53.

98. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:54.

99. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:55.

100. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:56.

101. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:57.

102. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:58.

103. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:59.

104. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:60.

```
<110> INCYTE GENOMICS, INC.
      YUE, Henry
      THORNTON, Michael
      RAMKUMAR, Jayalaxmi
      TANG, Y. Tom
      AZIMZAI, Yalda
      BAUGHN, Mariah R.
      YANG, Junming
      YAO, Monique G.
      LAL, Preeti
      WALIA, Narinder K.
      GANDHI, Ameena R.
      HAFALIA, April J.A.
      NGUYEN, Danniel B.
      PATTERSON, Chandra
      ELLIOTT, Vicki S.
      TRIBOULEY, Catherine M.
      LU, Dyung Aina M.
      XU, Yuming
      REDDY, Roopa
      HERNANDEZ, Roberto
      BOROWSKY, Mark L.
      LO, Terence P.
      LU, Yan
      POLICKY, Jennifer L.
      GREENE, Barrie D.
      SANJANWALA, Madhu S.
      RAUMANN, Brigette E.
      BURFORD, Neil
      ISON, Craig H.
      LEE, Ernestine A.
      DING, Li
      DAS, Debopriya
      KALLICK, Deborah A.
      KHAN, Farrah A.
      SEILHAMER, Jeffrey J.

<120> TRANSPORTERS AND ION CHANNELS

<130> PI-0183 PCT

<140> To Be Assigned
<141> Herewith

<150> 60/223,269; 60/224,456; 60/226,410; 60/228,140; 60/230,067;
      60/231,434
<151> 2000-08-03; 2000-08-10; 2000-08-18; 2000-08-25; 2000-08-31;
      2000-09-08

<160> 60
<170> PERL Program

<210> 1
<211> 308
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2194064CD1

<400> 1
Met Val Gly Gly Val Leu Ala Ser Leu Gly Phe Val Phe Ser Ala
```

1 5 10 15
Phe Ala Ser Asp Leu Leu His Leu Tyr Leu Gly Leu Gly Leu Leu
20 25 30
Ala Gly Phe Gly Trp Ala Leu Val Phe Ala Pro Ala Leu Gly Thr
35 40 45
Leu Ser Arg Tyr Phe Ser Arg Arg Arg Val Leu Ala Val Gly Leu
50 55 60
Ala Leu Thr Gly Asn Gly Ala Ser Ser Leu Leu Leu Ala Pro Ala
65 70 75
Leu Gln Leu Leu Leu Asp Thr Phe Gly Trp Arg Gly Ala Leu Leu
80 85 90
Leu Leu Gly Ala Ile Thr Leu His Leu Thr Pro Cys Gly Ala Leu
95 100 105
Leu Leu Pro Leu Val Leu Pro Gly Asp Pro Pro Ala Pro Pro Arg
110 115 120
Ser Pro Leu Ala Ala Leu Gly Leu Ser Leu Phe Thr Arg Arg Ala
125 130 135
Phe Ser Ile Phe Ala Leu Gly Thr Ala Leu Val Gly Gly Gly Tyr
140 145 150
Phe Val Pro Tyr Val His Leu Ala Pro His Ala Leu Asp Arg Gly
155 160 165
Leu Gly Gly Tyr Gly Ala Ala Leu Val Val Ala Val Ala Ala Met
170 175 180
Gly Asp Ala Gly Ala Arg Leu Val Cys Gly Trp Leu Ala Asp Gln
185 190 195
Gly Trp Val Pro Leu Pro Arg Leu Leu Ala Val Phe Gly Ala Leu
200 205 210
Thr Gly Leu Gly Leu Trp Val Val Gly Leu Val Pro Val Val Gly
215 220 225
Gly Glu Glu Ser Trp Gly Gly Pro Leu Leu Ala Ala Ala Val Ala
230 235 240
Tyr Gly Leu Ser Ala Gly Ser Tyr Ala Pro Leu Val Phe Gly Val
245 250 255
Leu Pro Gly Leu Val Gly Val Gly Gly Val Val Gln Ala Thr Gly
260 265 270
Leu Val Met Met Leu Met Ser Leu Gly Gly Leu Leu Gly Pro Pro
275 280 285
Leu Ser Gly Lys Asp Leu Ser Ser Gln Ile Cys Leu Gln Leu Ser
290 295 300
Ser Ala Pro Gly Val Arg Gly Phe
305

<210> 2
<211> 606
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2744094CD1

<400> 2
Met Ala Glu Gln Leu Ser Gln Gln Leu Pro Arg Thr Cys Leu Trp
1 5 10 15
His Leu Tyr Ile Thr Thr Val Ser Leu Pro Gly Tyr Met Val Ser
20 25 30
Cys Ile Ile Phe Phe Phe Val Val Pro Ile Val Phe Leu Thr Ile
35 40 45
Phe Ser Phe Trp Trp Leu Ser Tyr Trp Leu Glu Gln Gly Ser Gly
50 55 60
Thr Asn Ser Ser Arg Glu Ser Asn Gly Thr Met Ala Asp Leu Gly
65 70 75
Asn Ile Ala Asp Asn Pro Gln Leu Ser Phe Tyr Gln Leu Val Tyr

```
                80                  85                  90
Gly Leu Asn Ala Leu Leu Leu Ile Cys Val Gly Val Cys Ser Ser
                95                  100                 105
Gly Ile Phe Thr Lys Val Thr Arg Lys Ala Ser Thr Ala Leu His
                110                 115                 120
Asn Lys Leu Phe Asn Lys Val Phe Arg Cys Pro Met Ser Phe Phe
                125                 130                 135
Asp Thr Ile Pro Ile Gly Arg Leu Leu Asn Cys Phe Ala Gly Asp
                140                 145                 150
Leu Glu Gln Leu Asp Gln Leu Leu Pro Ile Phe Ser Glu Gln Phe
                155                 160                 165
Leu Val Leu Ser Leu Met Val Ile Ala Val Leu Leu Ile Val Ser
                170                 175                 180
Val Leu Ser Pro Tyr Ile Leu Leu Met Gly Ala Ile Ile Met Val
                185                 190                 195
Ile Cys Phe Ile Tyr Tyr Met Met Phe Lys Lys Ala Ile Gly Val
                200                 205                 210
Phe Lys Arg Leu Glu Asn Tyr Ser Arg Ser Pro Leu Phe Ser His
                215                 220                 225
Ile Leu Asn Ser Leu Gln Gly Leu Ser Ser Ile His Val Tyr Gly
                230                 235                 240
Lys Thr Glu Asp Phe Ile Ser Gln Phe Lys Arg Leu Thr Asp Ala
                245                 250                 255
Gln Asn Asn Tyr Leu Leu Leu Phe Leu Ser Ser Thr Arg Trp Met
                260                 265                 270
Ala Leu Arg Leu Glu Ile Met Thr Asn Leu Val Thr Leu Ala Val
                275                 280                 285
Ala Leu Phe Val Ala Phe Gly Ile Ser Ser Thr Pro Tyr Ser Phe
                290                 295                 300
Lys Val Met Ala Val Asn Ile Val Leu Gln Leu Ala Ser Ser Phe
                305                 310                 315
Gln Ala Thr Ala Arg Ile Gly Leu Glu Thr Glu Ala Gln Phe Thr
                320                 325                 330
Ala Val Glu Arg Ile Leu Gln Tyr Met Lys Met Cys Val Ser Glu
                335                 340                 345
Ala Pro Leu His Met Glu Gly Thr Ser Cys Pro Gln Gly Trp Pro
                350                 355                 360
Gln His Gly Glu Ile Ile Phe Gln Asp Tyr His Met Lys Tyr Arg
                365                 370                 375
Asp Asn Thr Pro Thr Val Leu His Gly Ile Asn Leu Thr Ile Arg
                380                 385                 390
Gly His Glu Val Val Gly Ile Val Gly Arg Thr Gly Ser Gly Lys
                395                 400                 405
Ser Ser Leu Gly Met Ala Leu Phe Arg Leu Val Glu Pro Met Ala
                410                 415                 420
Gly Arg Ile Leu Ile Asp Gly Val Asp Ile Cys Ser Ile Gly Leu
                425                 430                 435
Glu Asp Leu Arg Ser Lys Leu Ser Val Ile Pro Gln Asp Pro Val
                440                 445                 450
Leu Leu Ser Gly Thr Ile Arg Phe Asn Leu Asp Pro Phe Asp Arg
                455                 460                 465
His Thr Asp Gln Gln Ile Trp Asp Ala Leu Glu Arg Thr Phe Leu
                470                 475                 480
Thr Lys Ala Ile Ser Lys Phe Pro Lys Lys Leu His Thr Asp Val
                485                 490                 495
Val Glu Asn Gly Gly Tyr Phe Ser Val Gly Glu Arg Gln Leu Leu
                500                 505                 510
Cys Ile Ala Arg Ala Val Leu Arg Asn Ser Lys Ile Ile Leu Ile
                515                 520                 525
Asp Glu Ala Thr Ala Ser Ile Asp Met Glu Thr Asp Thr Leu Ile
                530                 535                 540
Gln Arg Thr Ile Arg Glu Ala Phe Gln Gly Cys Thr Val Leu Val
                545                 550                 555
```

```
Ile Ala His Arg Val Thr Thr Val Leu Asn Cys Asp Arg Ile Leu
                560                 565                 570
Val Met Gly Asn Gly Lys Val Val Glu Phe Asp Arg Pro Glu Val
                575                 580                 585
Leu Arg Lys Lys Pro Gly Ser Leu Phe Ala Ala Leu Met Ala Thr
                590                 595                 600
Ala Thr Ser Ser Leu Arg
                605
```

<210> 3
<211> 1642
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2798241CD1

<400> 3

```
Met Ser Thr Ala Ile Arg Glu Val Gly Val Trp Arg Gln Thr Arg
 1               5                  10                  15
Thr Leu Leu Leu Lys Asn Tyr Leu Ile Lys Cys Arg Thr Lys Lys
                20                  25                  30
Ser Ser Val Gln Glu Ile Leu Phe Pro Leu Phe Phe Leu Phe Trp
                35                  40                  45
Leu Ile Leu Ile Ser Met Met His Pro Asn Lys Lys Tyr Glu Glu
                50                  55                  60
Val Pro Asn Ile Glu Leu Asn Pro Met Asp Lys Phe Thr Leu Ser
                65                  70                  75
Asn Leu Ile Leu Gly Tyr Thr Pro Val Thr Asn Ile Thr Ser Ser
                80                  85                  90
Ile Met Gln Lys Val Ser Thr Asp His Leu Pro Asp Val Ile Ile
                95                 100                 105
Thr Glu Glu Tyr Thr Asn Glu Lys Glu Met Leu Thr Ser Ser Leu
                110                 115                 120
Ser Lys Pro Ser Asn Phe Val Gly Val Val Phe Lys Asp Ser Met
                125                 130                 135
Ser Tyr Glu Leu Arg Phe Phe Pro Asp Met Ile Pro Val Ser Ser
                140                 145                 150
Ile Tyr Met Asp Ser Arg Ala Gly Cys Ser Lys Ser Cys Glu Ala
                155                 160                 165
Ala Gln Tyr Trp Ser Ser Gly Phe Thr Val Leu Gln Ala Ser Ile
                170                 175                 180
Asp Ala Ala Ile Ile Gln Leu Lys Thr Asn Val Ser Leu Trp Lys
                185                 190                 195
Glu Leu Glu Ser Thr Lys Ala Val Ile Met Gly Glu Thr Ala Val
                200                 205                 210
Val Glu Ile Asp Thr Phe Pro Arg Gly Val Ile Leu Ile Tyr Leu
                215                 220                 225
Val Ile Ala Phe Ser Pro Phe Gly Tyr Phe Leu Ala Ile His Ile
                230                 235                 240
Val Ala Glu Lys Glu Lys Lys Ile Lys Glu Phe Leu Lys Ile Met
                245                 250                 255
Gly Leu His Asp Thr Ala Phe Trp Leu Ser Trp Val Leu Leu Tyr
                260                 265                 270
Thr Ser Leu Ile Phe Leu Met Ser Leu Leu Met Ala Val Ile Ala
                275                 280                 285
Thr Ala Ser Leu Leu Phe Pro Gln Ser Ser Ser Ile Val Ile Phe
                290                 295                 300
Leu Leu Phe Phe Leu Tyr Gly Leu Ser Ser Val Phe Phe Ala Leu
                305                 310                 315
Met Leu Thr Pro Leu Phe Lys Lys Ser Lys His Val Gly Ile Val
                320                 325                 330
```

Glu Phe Phe Val Thr Val Ala Phe Gly Phe Ile Gly Leu Met Ile
335 340 345
Ile Leu Ile Glu Ser Phe Pro Lys Ser Leu Val Trp Leu Phe Ser
350 355 360
Pro Phe Cys His Cys Thr Phe Val Ile Gly Ile Ala Gln Val Met
365 370 375
His Leu Glu Asp Phe Asn Glu Gly Ala Ser Phe Ser Asn Leu Thr
380 385 390
Ala Gly Pro Tyr Pro Leu Ile Ile Thr Ile Ile Met Leu Thr Leu
395 400 405
Asn Ser Ile Phe Tyr Val Leu Leu Ala Val Tyr Leu Asp Gln Val
410 415 420
Ile Pro Gly Glu Phe Gly Leu Arg Arg Ser Ser Leu Tyr Phe Leu
425 430 435
Lys Pro Ser Tyr Trp Ser Lys Ser Lys Arg Asn Tyr Glu Glu Leu
440 445 450
Ser Glu Gly Asn Val Asn Gly Asn Ile Ser Phe Ser Glu Ile Ile
455 460 465
Glu Pro Val Ser Ser Glu Phe Val Gly Lys Glu Ala Ile Arg Ile
470 475 480
Ser Gly Ile Gln Lys Thr Tyr Arg Lys Lys Gly Glu Asn Val Glu
485 490 495
Ala Leu Arg Asn Leu Ser Phe Asp Ile Tyr Glu Gly Gln Ile Thr
500 505 510
Ala Leu Leu Gly His Ser Gly Thr Gly Lys Ser Thr Leu Met Asn
515 520 525
Ile Leu Cys Gly Leu Cys Pro Pro Ser Asp Gly Phe Ala Ser Ile
530 535 540
Tyr Gly His Arg Val Ser Glu Ile Asp Glu Met Phe Glu Ala Arg
545 550 555
Lys Met Ile Gly Ile Cys Pro Gln Leu Asp Ile His Phe Asp Val
560 565 570
Leu Thr Val Glu Glu Asn Leu Ser Ile Leu Ala Ser Ile Lys Gly
575 580 585
Ile Pro Ala Asn Asn Ile Ile Gln Glu Val Gln Lys Val Leu Leu
590 595 600
Asp Leu Asp Met Gln Thr Ile Lys Asp Asn Gln Ala Lys Lys Leu
605 610 615
Ser Gly Gly Gln Lys Arg Lys Leu Ser Leu Gly Ile Ala Val Leu
620 625 630
Gly Asn Pro Lys Ile Leu Leu Leu Asp Glu Pro Thr Ala Gly Met
635 640 645
Asp Pro Cys Ser Arg His Ile Val Trp Asn Leu Leu Lys Tyr Arg
650 655 660
Lys Ala Asn Arg Val Thr Val Phe Ser Thr His Phe Met Asp Glu
665 670 675
Ala Asp Ile Leu Ala Asp Arg Lys Ala Val Ile Ser Gln Gly Met
680 685 690
Leu Lys Cys Val Gly Ser Ser Met Phe Leu Lys Ser Lys Trp Gly
695 700 705
Ile Gly Tyr Arg Leu Ser Met Tyr Ile Asp Lys Tyr Cys Ala Thr
710 715 720
Glu Ser Leu Ser Ser Leu Val Lys Gln His Ile Pro Gly Ala Thr
725 730 735
Leu Leu Gln Gln Asn Asp Gln Gln Leu Val Tyr Ser Leu Pro Phe
740 745 750
Lys Asp Met Asp Lys Phe Ser Gly Leu Phe Ser Ala Leu Asp Ser
755 760 765
His Ser Asn Leu Gly Val Ile Ser Tyr Gly Val Ser Met Thr Thr
770 775 780
Leu Glu Asp Val Phe Leu Lys Leu Glu Val Glu Ala Glu Ile Asp
785 790 795
Gln Ala Asp Tyr Ser Val Phe Thr Gln Gln Pro Leu Glu Glu Glu

```
                800                 805                 810
Met Asp Ser Lys Ser Phe Asp Glu Met Glu Gln Ser Leu Leu Ile
                815                 820                 825
Leu Ser Glu Thr Lys Ala Ser Leu Val Ser Thr Met Ser Leu Trp
                830                 835                 840
Lys Gln Gln Met Tyr Thr Ile Ala Lys Phe His Phe Phe Thr Leu
                845                 850                 855
Lys Arg Glu Ser Lys Ser Val Arg Ser Val Leu Leu Leu Leu Leu
                860                 865                 870
Ile Phe Phe Thr Val Gln Ile Phe Met Phe Leu Val His His Ser
                875                 880                 885
Phe Lys Asn Ala Val Val Pro Ile Lys Leu Val Pro Asp Leu Tyr
                890                 895                 900
Phe Leu Lys Pro Gly Asp Lys Pro His Lys Tyr Lys Thr Ser Leu
                905                 910                 915
Leu Leu Gln Asn Ser Ala Asp Ser Asp Ile Ser Asp Leu Ile Ser
                920                 925                 930
Phe Phe Thr Ser Gln Asn Ile Met Val Thr Met Ile Asn Asp Ser
                935                 940                 945
Asp Tyr Val Ser Val Ala Pro His Ser Ala Ala Leu Asn Val Val
                950                 955                 960
His Ser Glu Lys Asp Tyr Val Phe Ala Ala Val Phe Asn Ser Thr
                965                 970                 975
Met Val Tyr Ser Leu Pro Ile Leu Val Asn Ile Ile Ser Asn Tyr
                980                 985                 990
Tyr Leu Tyr His Leu Asn Val Thr Glu Thr Ile Gln Ile Trp Ser
                995                1000                1005
Thr Pro Phe Phe Gln Glu Ile Thr Asp Ile Val Phe Lys Ile Glu
               1010                1015                1020
Leu Tyr Phe Gln Ala Ala Leu Leu Gly Ile Ile Val Thr Ala Met
               1025                1030                1035
Pro Pro Tyr Phe Ala Met Glu Asn Ala Glu Asn His Lys Ile Lys
               1040                1045                1050
Ala Tyr Thr Gln Leu Lys Leu Ser Gly Leu Leu Pro Ser Ala Tyr
               1055                1060                1065
Trp Ile Gly Gln Ala Val Val Asp Ile Pro Leu Phe Phe Ile Ile
               1070                1075                1080
Leu Ile Leu Met Leu Gly Ser Leu Leu Ala Phe His Tyr Gly Leu
               1085                1090                1095
Tyr Phe Tyr Thr Val Lys Phe Leu Ala Val Val Phe Cys Leu Ile
               1100                1105                1110
Gly Tyr Val Pro Ser Val Ile Leu Phe Thr Tyr Ile Ala Ser Phe
               1115                1120                1125
Thr Phe Lys Lys Ile Leu Asn Thr Lys Glu Phe Trp Ser Phe Ile
               1130                1135                1140
Tyr Ser Val Ala Ala Leu Ala Cys Ile Ala Ile Thr Glu Ile Thr
               1145                1150                1155
Phe Phe Met Gly Tyr Thr Ile Ala Thr Ile Leu His Tyr Ala Phe
               1160                1165                1170
Cys Ile Ile Ile Pro Ile Tyr Pro Leu Leu Gly Cys Leu Ile Ser
               1175                1180                1185
Phe Ile Lys Ile Ser Trp Lys Asn Val Arg Lys Asn Val Asp Thr
               1190                1195                1200
Tyr Asn Pro Trp Asp Arg Leu Ser Val Ala Val Ile Ser Pro Tyr
               1205                1210                1215
Leu Gln Cys Val Leu Trp Ile Phe Leu Leu Gln Tyr Tyr Glu Lys
               1220                1225                1230
Lys Tyr Gly Gly Arg Ser Ile Arg Lys Asp Pro Phe Phe Arg Asn
               1235                1240                1245
Leu Ser Thr Lys Ser Lys Asn Arg Lys Leu Pro Glu Pro Pro Asp
               1250                1255                1260
Asn Glu Asp Glu Asp Glu Asp Val Lys Ala Glu Arg Leu Lys Val
               1265                1270                1275
```

Lys Glu Leu Met Gly Cys Gln Cys Cys Glu Glu Lys Pro Ser Ile
1280 1285 1290
Met Val Ser Asn Leu His Lys Glu Tyr Asp Asp Lys Lys Asp Phe
1295 1300 1305
Leu Leu Ser Arg Lys Val Lys Lys Val Ala Thr Lys Tyr Ile Ser
1310 1315 1320
Phe Cys Val Lys Lys Gly Glu Ile Leu Gly Leu Leu Gly Pro Asn
1325 1330 1335
Gly Ala Gly Lys Ser Thr Ile Ile Asn Ile Leu Val Gly Asp Ile
1340 1345 1350
Glu Pro Thr Ser Gly Gln Val Phe Leu Gly Asp Tyr Ser Ser Glu
1355 1360 1365
Thr Ser Glu Asp Asp Asp Ser Leu Lys Cys Met Gly Tyr Cys Pro
1370 1375 1380
Gln Ile Asn Pro Leu Trp Pro Asp Thr Thr Leu Gln Glu His Phe
1385 1390 1395
Glu Ile Tyr Gly Ala Val Lys Gly Met Ser Ala Ser Asp Met Lys
1400 1405 1410
Glu Val Ile Ser Arg Ile Thr His Ala Leu Asp Leu Lys Glu His
1415 1420 1425
Leu Gln Lys Thr Val Lys Lys Leu Pro Ala Gly Ile Lys Arg Lys
1430 1435 1440
Leu Cys Phe Ala Leu Ser Met Leu Gly Asn Pro Gln Ile Thr Leu
1445 1450 1455
Leu Asp Glu Pro Ser Thr Gly Met Asp Pro Lys Ala Lys Gln His
1460 1465 1470
Met Trp Arg Ala Ile Arg Thr Ala Phe Lys Asn Arg Lys Arg Ala
1475 1480 1485
Ala Ile Leu Thr Thr His Tyr Met Glu Glu Ala Glu Ala Val Cys
1490 1495 1500
Asp Arg Val Ala Ile Met Val Ser Gly Gln Leu Arg Cys Ile Gly
1505 1510 1515
Thr Val Gln His Leu Lys Ser Lys Phe Gly Lys Gly Tyr Phe Leu
1520 1525 1530
Glu Ile Lys Leu Lys Asp Trp Ile Glu Asn Leu Glu Val Asp Arg
1535 1540 1545
Leu Gln Arg Glu Ile Gln Tyr Ile Phe Pro Asn Ala Ser Arg Gln
1550 1555 1560
Glu Ser Phe Ser Ser Ile Leu Ala Tyr Lys Ile Pro Lys Glu Asp
1565 1570 1575
Val Gln Ser Leu Ser Gln Ser Phe Phe Lys Leu Glu Glu Ala Lys
1580 1585 1590
His Ala Phe Ala Ile Glu Glu Tyr Ser Phe Ser Gln Ala Thr Leu
1595 1600 1605
Glu Gln Val Phe Val Glu Leu Thr Lys Glu Gln Glu Glu Glu Asp
1610 1615 1620
Asn Ser Cys Gly Thr Leu Asn Ser Thr Leu Trp Trp Glu Arg Thr
1625 1630 1635
Gln Glu Asp Arg Val Val Phe
1640

<210> 4
<211> 659
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 3105257CD1

<400> 4
Met Gly Arg Gly Ala Gly Ala Ala Leu Gly Arg Trp Ser Arg Ala
1 5 10 15

Pro Leu Glu Glu Leu Leu Pro Gly Arg Gly Ser Gly Arg Leu Gly
20 25 30
Gly Pro Arg Gly Pro Arg Thr Ala Pro Gly Ala Val Gly Leu Gly
35 40 45
Pro Ala Ala Ala Gly Glu Glu Ala Trp Arg Arg Gly Arg Ala Ala
50 55 60
Pro Ser Arg Asp Asp Gln Arg Leu Arg Pro Met Ala Pro Gly Leu
65 70 75
Ser Glu Ala Gly Lys Leu Leu Gly Leu Glu Tyr Pro Glu Arg Gln
80 85 90
Arg Leu Ala Ala Ala Val Gly Phe Leu Thr Met Ser Gly Val Ile
95 100 105
Ser Met Ser Ala Pro Phe Phe Leu Gly Lys Ile Ile Asp Ala Ile
110 115 120
Tyr Thr Asn Pro Thr Val Asp Tyr Ser Asp Asn Leu Thr Arg Leu
125 130 135
Cys Leu Gly Leu Ser Ala Val Phe Leu Cys Gly Ala Ala Ala Asn
140 145 150
Ala Ile Arg Val Tyr Leu Met Gln Thr Ser Gly Gln Arg Ile Val
155 160 165
Asn Arg Leu Arg Thr Ser Leu Phe Ser Ser Ile Leu Arg Gln Glu
170 175 180
Val Ala Phe Phe Asp Lys Thr Arg Thr Gly Glu Leu Ile Asn Arg
185 190 195
Leu Ser Ser Asp Thr Ala Leu Leu Gly Arg Ser Val Thr Glu Asn
200 205 210
Leu Ser Asp Gly Leu Arg Ala Gly Ala Gln Ala Ser Val Gly Ile
215 220 225
Ser Met Met Phe Phe Val Ser Pro Asn Leu Ala Thr Phe Val Leu
230 235 240
Ser Val Val Pro Pro Val Ser Ile Ile Ala Val Ile Tyr Gly Arg
245 250 255
Tyr Leu Arg Lys Leu Thr Lys Val Thr Gln Asp Ser Leu Ala Gln
260 265 270
Ala Thr Gln Leu Ala Glu Glu Arg Ile Gly Asn Val Arg Thr Val
275 280 285
Arg Ala Phe Gly Lys Glu Met Thr Glu Ile Glu Lys Tyr Ala Ser
290 295 300
Lys Val Asp His Val Met Gln Leu Ala Arg Lys Glu Ala Phe Ala
305 310 315
Arg Ala Gly Phe Phe Gly Ala Thr Gly Leu Ser Gly Asn Leu Ile
320 325 330
Val Leu Ser Val Leu Tyr Lys Gly Gly Leu Leu Met Gly Ser Ala
335 340 345
His Met Thr Val Gly Glu Leu Ser Ser Phe Leu Met Tyr Ala Phe
350 355 360
Trp Val Gly Ile Ser Ile Gly Gly Leu Ser Ser Phe Tyr Ser Glu
365 370 375
Leu Met Lys Gly Leu Gly Ala Gly Gly Arg Leu Trp Glu Leu Leu
380 385 390
Glu Arg Glu Pro Lys Leu Pro Phe Asn Glu Gly Val Ile Leu Asn
395 400 405
Glu Lys Ser Phe Gln Gly Ala Leu Glu Phe Lys Asn Val His Phe
410 415 420
Ala Tyr Pro Ala Arg Pro Glu Val Pro Ile Phe Gln Asp Phe Ser
425 430 435
Leu Ser Ile Pro Ser Gly Ser Val Thr Ala Leu Val Gly Pro Ser
440 445 450
Gly Ser Gly Lys Ser Thr Val Leu Ser Leu Leu Leu Arg Leu Tyr
455 460 465
Asp Pro Ala Ser Gly Thr Ile Ser Leu Asp Gly His Asp Ile Arg
470 475 480
Gln Leu Asn Pro Val Trp Leu Arg Ser Lys Ile Gly Thr Val Ser

```
                485                 490                 495
Gln Glu Pro Ile Leu Phe Ser Cys Ser Ile Ala Glu Asn Ile Ala
                500                 505                 510
Tyr Gly Ala Asp Asp Pro Ser Ser Val Thr Ala Glu Glu Ile Gln
                515                 520                 525
Arg Val Ala Glu Val Ala Asn Thr Val Ala Phe Ile Arg Asn Phe
                530                 535                 540
Pro Gln Gly Phe Asn Thr Val Val Gly Glu Lys Gly Val Leu Leu
                545                 550                 555
Ser Gly Gly Gln Lys Gln Arg Ile Ala Ile Ala Arg Ala Leu Leu
                560                 565                 570
Lys Asn Pro Lys Ile Leu Leu Leu Asp Glu Ala Thr Ser Ala Leu
                575                 580                 585
Asp Ala Glu Asn Glu Tyr Leu Val Gln Glu Ala Leu Asp Arg Leu
                590                 595                 600
Met Asp Gly Arg Thr Val Leu Val Ile Ala His Arg Leu Ser Thr
                605                 610                 615
Ile Lys Asn Ala Asn Met Val Ala Val Leu Asp Gln Gly Lys Ile
                620                 625                 630
Thr Glu Tyr Gly Lys His Glu Glu Leu Leu Ser Lys Pro Asn Gly
                635                 640                 645
Ile Tyr Arg Lys Leu Met Asn Lys Gln Ser Phe Ile Ser Ala
                650                 655

<210> 5
<211> 1592
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 3200979CD1

<400> 5
Met Val Lys Lys Glu Ile Ser Val Arg Gln Gln Ile Gln Ala Leu
  1             5                   10                  15
Leu Tyr Lys Asn Phe Leu Lys Lys Trp Arg Ile Lys Arg Glu Phe
                20                  25                  30
Ile Gly Leu Tyr Leu Cys Ile Phe Ser Glu His Phe Arg Ala Thr
                35                  40                  45
Arg Phe Pro Glu Gln Pro Pro Lys Val Leu Gly Ser Val Asp Gln
                50                  55                  60
Phe Asn Asp Ser Gly Leu Val Val Ala Tyr Thr Pro Val Ser Asn
                65                  70                  75
Ile Thr Gln Arg Ile Met Asn Lys Met Ala Leu Ala Ser Phe Met
                80                  85                  90
Lys Gly Arg Thr Val Ile Gly Thr Pro Asp Glu Glu Thr Met Asp
                95                  100                 105
Ile Glu Leu Pro Lys Lys Tyr His Glu Met Val Gly Val Ile Phe
                110                 115                 120
Ser Asp Thr Phe Ser Tyr Arg Leu Lys Phe Asn Trp Gly Tyr Arg
                125                 130                 135
Ile Pro Val Ile Lys Glu His Ser Glu Tyr Thr Glu His Cys Trp
                140                 145                 150
Ala Met His Gly Glu Ile Phe Cys Tyr Leu Ala Lys Tyr Trp Leu
                155                 160                 165
Lys Gly Phe Val Ala Phe Gln Ala Ala Ile Asn Ala Ala Ile Ile
                170                 175                 180
Glu Val Thr Thr Asn His Ser Val Met Glu Glu Leu Thr Ser Val
                185                 190                 195
Ile Gly Ile Asn Met Lys Ile Pro Pro Phe Ile Ser Lys Gly Glu
                200                 205                 210
Ile Met Asn Glu Trp Phe His Phe Thr Cys Leu Val Ser Phe Ser
```

```
                    215                      220                      225
Ser Phe Ile Tyr Phe Ala Ser Leu Asn Val Ala Arg Glu Arg Gly
                    230                      235                      240
Lys Phe Lys Lys Leu Met Thr Val Met Gly Leu Arg Glu Ser Ala
                    245                      250                      255
Phe Trp Leu Ser Trp Gly Leu Thr Tyr Ile Cys Phe Ile Phe Ile
                    260                      265                      270
Met Ser Ile Phe Met Ala Leu Val Ile Thr Ser Ile Pro Ile Val
                    275                      280                      285
Phe His Thr Gly Phe Met Val Ile Phe Thr Leu Tyr Ser Leu Tyr
                    290                      295                      300
Gly Leu Ser Leu Ile Ala Leu Ala Phe Leu Met Ser Val Leu Ile
                    305                      310                      315
Arg Lys Pro Met Leu Ala Gly Leu Ala Gly Phe Leu Phe Thr Val
                    320                      325                      330
Phe Trp Gly Cys Leu Gly Phe Thr Val Leu Tyr Arg Gln Leu Pro
                    335                      340                      345
Leu Ser Leu Gly Trp Val Leu Ser Leu Leu Ser Pro Phe Ala Phe
                    350                      355                      360
Thr Ala Gly Met Ala Gln Ile Thr His Leu Asp Asn Tyr Leu Ser
                    365                      370                      375
Gly Val Ile Phe Pro Asp Pro Ser Gly Asp Ser Tyr Lys Met Ile
                    380                      385                      390
Ala Thr Phe Phe Ile Leu Ala Phe Asp Thr Leu Phe Tyr Leu Ile
                    395                      400                      405
Phe Thr Leu Tyr Phe Glu Arg Val Leu Pro Asp Lys Asp Gly His
                    410                      415                      420
Gly Asp Ser Pro Leu Phe Phe Leu Lys Ser Ser Phe Trp Ser Lys
                    425                      430                      435
His Gln Asn Thr His His Glu Ile Phe Glu Asn Glu Ile Asn Pro
                    440                      445                      450
Glu His Ser Ser Asp Asp Ser Phe Glu Pro Val Ser Pro Glu Phe
                    455                      460                      465
His Gly Lys Glu Ala Ile Arg Ile Arg Asn Val Ile Lys Glu Tyr
                    470                      475                      480
Asn Gly Lys Thr Gly Lys Val Glu Ala Leu Gln Gly Ile Phe Phe
                    485                      490                      495
Asp Ile Tyr Glu Gly Gln Ile Thr Ala Ile Leu Gly His Asn Gly
                    500                      505                      510
Ala Gly Lys Ser Thr Leu Leu Asn Ile Leu Ser Gly Leu Ser Val
                    515                      520                      525
Ser Thr Glu Gly Ser Ala Thr Ile Tyr Asn Thr Gln Leu Ser Glu
                    530                      535                      540
Ile Thr Asp Met Glu Glu Ile Arg Lys Asn Ile Gly Phe Cys Pro
                    545                      550                      555
Gln Phe Asn Phe Gln Phe Asp Phe Leu Thr Val Arg Glu Asn Leu
                    560                      565                      570
Arg Val Phe Ala Lys Ile Lys Gly Ile Gln Pro Lys Glu Val Glu
                    575                      580                      585
Gln Glu Val Leu Leu Leu Asp Glu Pro Thr Ala Gly Leu Asp Pro
                    590                      595                      600
Phe Ser Arg His Arg Val Trp Ser Leu Leu Lys Glu His Lys Val
                    605                      610                      615
Asp Arg Leu Ile Leu Phe Ser Thr Gln Phe Met Asp Glu Ala Asp
                    620                      625                      630
Ile Leu Ala Asp Arg Lys Val Phe Leu Ser Asn Gly Lys Leu Lys
                    635                      640                      645
Cys Ala Gly Ser Ser Leu Phe Leu Lys Arg Lys Trp Gly Ile Gly
                    650                      655                      660
Tyr His Leu Ser Leu His Arg Asn Glu Met Cys Asp Thr Glu Lys
                    665                      670                      675
Ile Thr Ser Leu Ile Lys Gln His Ile Pro Asp Ala Lys Leu Thr
                    680                      685                      690
```

Thr Glu Ser Glu Glu Lys Leu Val Tyr Ser Leu Pro Leu Glu Lys
695 700 705
Thr Asn Lys Phe Pro Asp Leu Tyr Ser Asp Leu Asp Lys Cys Ser
710 715 720
Asp Gln Gly Ile Arg Asn Tyr Ala Val Ser Val Thr Ser Leu Asn
725 730 735
Glu Val Phe Leu Asn Leu Glu Gly Lys Ser Ala Ile Asp Glu Pro
740 745 750
Asp Phe Asp Ile Gly Lys Gln Glu Lys Ile His Val Thr Arg Asn
755 760 765
Thr Gly Asp Glu Ser Glu Met Glu Gln Val Leu Cys Ser Leu Pro
770 775 780
Glu Thr Arg Lys Ala Val Ser Ser Ala Ala Leu Trp Arg Arg Gln
785 790 795
Ile Tyr Ala Val Ala Thr Leu Arg Phe Leu Lys Leu Arg Arg Glu
800 805 810
Arg Arg Ala Leu Leu Cys Leu Leu Leu Val Leu Gly Ile Ala Phe
815 820 825
Ile Pro Ile Ile Leu Glu Lys Ile Met Tyr Lys Val Thr Arg Glu
830 835 840
Thr His Cys Trp Glu Phe Ser Pro Ser Met Tyr Phe Leu Ser Leu
845 850 855
Glu Gln Ile Pro Lys Thr Pro Leu Thr Ser Leu Leu Ile Val Asn
860 865 870
Asn Thr Gly Ser Asn Ile Glu Asp Leu Val His Ser Leu Lys Cys
875 880 885
Gln Asp Ile Val Leu Glu Ile Asp Asp Phe Arg Asn Arg Asn Gly
890 895 900
Ser Asp Asp Pro Ser Tyr Asn Gly Ala Ile Ile Val Ser Gly Asp
905 910 915
Gln Lys Asp Tyr Arg Phe Ser Val Ala Cys Asn Thr Lys Lys Ser
920 925 930
Asn Cys Phe Pro Val Leu Met Gly Ile Val Ser Asn Ala Leu Ile
935 940 945
Gly Ile Phe Asn Phe Thr Glu Leu Ile Gln Met Glu Ser Thr Ser
950 955 960
Phe Phe Arg Asp Asp Ile Val Leu Asp Leu Gly Phe Ile Asp Gly
965 970 975
Ser Ile Phe Leu Leu Leu Ile Thr Asn Cys Ile Ser Pro Tyr Ile
980 985 990
Gly Ile Ser Ser Ile Ser Asp Tyr Lys Ile Pro Ser Ser Ile Pro
995 1000 1005
Ser Ile Leu Cys Gln Lys Asn Val Gln Ser Gln Leu Trp Ile Ser
1010 1015 1020
Gly Leu Trp Pro Ser Ala Tyr Trp Cys Gly Gln Ala Leu Val Asp
1025 1030 1035
Ile Pro Leu His Phe Leu Ile Leu Leu Ser Ile His Leu Ile Tyr
1040 1045 1050
Tyr Phe Ser Phe Leu Gly Phe Gln Leu Pro Trp Glu Leu Met Phe
1055 1060 1065
Val Leu Val Val Cys Ile Ile Gly Cys Ala Ala Ser Leu Ile Phe
1070 1075 1080
Leu Met Tyr Val Leu Ser Phe Ile Phe Cys Lys Trp Arg Lys Asn
1085 1090 1095
Asn Gly Phe Trp Ser Phe Gly Phe Phe Ile Val Leu Ile Cys Val
1100 1105 1110
Ser Thr Ile Leu Val Ser Thr Lys Tyr Glu Lys Pro Asn Leu Ile
1115 1120 1125
Leu Cys Met Ile Phe Ile Pro Ser Phe Thr Phe Leu Asp Met Ser
1130 1135 1140
Leu Leu Ile Gln Leu Asn Phe Met Tyr Met Arg Asn Leu Asp Ser
1145 1150 1155
Leu Asp Asn Arg Ile Asn Glu Val Asn Lys Thr Ile Leu Leu Thr

1160 1165 1170
Asn Leu Ile Pro Tyr Leu Gln Ser Val Ile Phe Leu Phe Val Ile
1175 1180 1185
Arg Cys Leu Glu Met Lys Tyr Gly Asn Glu Ile Met Asn Lys Asp
1190 1195 1200
Pro Val Phe Arg Ile Ser Pro Arg Ser Arg Gly Thr His Thr Asn
1205 1210 1215
Pro Glu Glu Pro Glu Glu Asp Val Gln Ala Glu Arg Val Gln Ala
1220 1225 1230
Ala Asn Ala Leu Thr Thr Pro Asn Leu Glu Glu Glu Pro Val Ile
1235 1240 1245
Thr Ala Ser Cys Leu His Lys Glu Tyr Tyr Glu Thr Lys Lys Ser
1250 1255 1260
Cys Phe Ser Thr Thr Lys Lys Lys Ala Ala Ile Arg Asn Val Ser
1265 1270 1275
Phe Cys Val Lys Lys Gly Glu Val Leu Gly Leu Leu Gly His Asn
1280 1285 1290
Gly Ala Gly Lys Ser Thr Ser Ile Lys Met Ile Thr Gly Cys Thr
1295 1300 1305
Val Pro Thr Ala Gly Val Val Val Leu Gln Gly Asn Arg Ala Ser
1310 1315 1320
Val Arg Gln Gln Arg Asp Asn Ser Leu Lys Phe Leu Gly Tyr Cys
1325 1330 1335
Pro Gln Glu Asn Ser Leu Trp Pro Lys Leu Thr Met Lys Glu His
1340 1345 1350
Leu Glu Leu Tyr Ala Ala Val Lys Gly Leu Gly Lys Glu Asp Ala
1355 1360 1365
Ala Leu Ser Ile Ser Arg Leu Val Glu Ala Leu Lys Leu Gln Glu
1370 1375 1380
Gln Leu Lys Ala Pro Val Lys Thr Leu Ser Glu Gly Ile Lys Arg
1385 1390 1395
Lys Leu Cys Phe Val Leu Ser Ile Leu Gly Asn Pro Ser Val Val
1400 1405 1410
Leu Leu Asp Glu Pro Phe Thr Gly Met Asp Pro Glu Gly Gln Gln
1415 1420 1425
Gln Met Trp Gln Ile Leu Gln Ala Thr Ile Lys Asn Gln Glu Arg
1430 1435 1440
Gly Thr Leu Leu Thr Thr His Tyr Met Ser Glu Ala Lys Ser Leu
1445 1450 1455
Cys Asp Arg Val Ala Ile Met Val Ser Gly Thr Leu Arg Cys Ile
1460 1465 1470
Gly Ser Ile Gln His Leu Lys Asn Lys Phe Gly Lys Asp Tyr Leu
1475 1480 1485
Leu Glu Ile Lys Met Lys Glu Pro Thr Gln Val Glu Ala Leu His
1490 1495 1500
Thr Glu Ile Leu Lys Leu Phe Pro Gln Ala Ala Trp Gln Glu Arg
1505 1510 1515
Tyr Ser Ser Leu Met Ala Tyr Lys Leu Pro Val Glu Asp Val His
1520 1525 1530
Pro Leu Ser Arg Ala Phe Phe Lys Leu Glu Ala Met Lys Gln Thr
1535 1540 1545
Phe Asn Leu Glu Glu Tyr Ser Leu Ser Gln Ala Thr Leu Glu Gln
1550 1555 1560
Val Phe Leu Glu Leu Cys Lys Glu Gln Glu Leu Gly Asn Val Asp
1565 1570 1575
Asp Lys Ile Asp Thr Thr Val Glu Trp Lys Leu Leu Pro Gln Glu
1580 1585 1590
Asp Pro

<210> 6
<211> 382
<212> PRT

<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 6754139CD1

<400> 6
```
Met Asp Glu Arg Asn Gln Val Leu Thr Leu Tyr Leu Trp Ile Arg
 1               5                  10                  15
Gln Glu Trp Thr Asp Ala Tyr Leu Arg Trp Asp Pro Asn Ala Tyr
                20                  25                  30
Gly Gly Leu Asp Ala Ile Arg Ile Pro Ser Ser Leu Val Trp Arg
                35                  40                  45
Pro Asp Ile Val Leu Tyr Asn Lys Ala Asp Ala Gln Pro Pro Gly
                50                  55                  60
Ser Ala Ser Thr Asn Val Val Leu Arg His Asp Gly Ala Val Arg
                65                  70                  75
Trp Asp Ala Pro Ala Ile Thr Arg Ser Ser Cys Arg Val Asp Val
                80                  85                  90
Ala Ala Phe Pro Phe Asp Ala Gln His Cys Gly Leu Thr Phe Gly
                95                 100                 105
Ser Trp Thr His Gly Gly His Gln Val Asp Val Arg Pro Arg Gly
               110                 115                 120
Ala Ala Ala Ser Leu Ala Asp Phe Val Glu Asn Val Glu Trp Arg
               125                 130                 135
Val Leu Gly Met Pro Ala Arg Arg Arg Val Leu Thr Tyr Gly Cys
               140                 145                 150
Cys Ser Glu Pro Tyr Pro Asp Val Thr Phe Thr Leu Leu Leu Arg
               155                 160                 165
Arg Arg Ala Ala Ala Tyr Val Cys Asn Leu Leu Leu Pro Cys Val
               170                 175                 180
Leu Ile Ser Leu Leu Ala Pro Leu Ala Phe His Leu Pro Ala Asp
               185                 190                 195
Ser Gly Glu Lys Val Ser Leu Gly Val Thr Val Leu Leu Ala Leu
               200                 205                 210
Thr Val Phe Gln Leu Leu Leu Ala Glu Ser Met Pro Pro Ala Glu
               215                 220                 225
Ser Val Pro Leu Ile Gly Lys Tyr Tyr Met Ala Thr Met Thr Met
               230                 235                 240
Val Thr Phe Ser Thr Ala Leu Thr Ile Leu Ile Met Asn Leu His
               245                 250                 255
Tyr Cys Gly Pro Ser Val Arg Pro Val Pro Ala Trp Ala Arg Ala
               260                 265                 270
Leu Leu Leu Gly His Leu Ala Arg Gly Leu Cys Val Arg Glu Arg
               275                 280                 285
Gly Glu Pro Cys Gly Gln Ser Arg Pro Pro Glu Leu Ser Pro Ser
               290                 295                 300
Pro Gln Ser Pro Glu Gly Gly Ala Gly Pro Pro Ala Gly Pro Cys
               305                 310                 315
His Glu Pro Arg Cys Leu Cys Arg Gln Glu Ala Leu Leu His His
               320                 325                 330
Val Ala Thr Ile Ala Asn Thr Phe Arg Ser His Arg Ala Ala Gln
               335                 340                 345
Arg Cys His Glu Asp Trp Lys Arg Leu Ala Arg Val Met Asp Arg
               350                 355                 360
Phe Phe Leu Ala Ile Phe Phe Ser Met Ala Leu Val Met Ser Leu
               365                 370                 375
Leu Val Leu Val Gln Ala Leu
               380
```

<210> 7
<211> 1115
<212> PRT

<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 6996659CD1

<400> 7
Met Arg Arg Leu Ser Leu Trp Trp Leu Leu Ser Arg Val Cys Leu
1 5 10 15
Leu Leu Pro Pro Pro Cys Ala Leu Val Leu Ala Gly Val Pro Ser
20 25 30
Ser Ser Ser His Pro Gln Pro Cys Gln Ile Leu Lys Arg Ile Gly
35 40 45
His Ala Val Arg Val Gly Ala Val His Leu Gln Pro Trp Thr Thr
50 55 60
Ala Pro Arg Ala Ala Ser Arg Ala Pro Asp Asp Ser Arg Ala Gly
65 70 75
Ala Gln Arg Asp Glu Pro Glu Pro Gly Thr Arg Arg Ser Pro Ala
80 85 90
Pro Ser Pro Gly Ala Arg Trp Leu Gly Ser Thr Leu His Gly Arg
95 100 105
Gly Pro Pro Gly Ser Arg Lys Pro Gly Glu Gly Ala Arg Ala Glu
110 115 120
Ala Leu Trp Pro Arg Asp Ala Leu Leu Phe Ala Val Asp Asn Leu
125 130 135
Asn Arg Val Glu Gly Leu Leu Pro Tyr Asn Leu Ser Leu Glu Val
140 145 150
Val Met Ala Ile Glu Ala Gly Leu Gly Asp Leu Pro Leu Leu Pro
155 160 165
Phe Ser Ser Pro Ser Ser Pro Trp Ser Ser Asp Pro Phe Ser Phe
170 175 180
Leu Gln Ser Val Cys His Thr Val Val Val Gln Gly Val Ser Ala
185 190 195
Leu Leu Ala Phe Pro Gln Ser Gln Gly Glu Met Met Glu Leu Asp
200 205 210
Leu Val Ser Leu Val Leu His Ile Pro Val Ile Ser Ile Val Arg
215 220 225
His Glu Phe Pro Arg Glu Ser Gln Asn Pro Leu His Leu Gln Leu
230 235 240
Ser Leu Glu Asn Ser Leu Ser Ser Asp Ala Asp Val Thr Val Ser
245 250 255
Ile Leu Thr Met Asn Asn Trp Tyr Asn Phe Ser Leu Leu Leu Cys
260 265 270
Gln Glu Asp Trp Asn Ile Thr Asp Phe Leu Leu Leu Thr Gln Asn
275 280 285
Asn Ser Lys Phe His Leu Gly Ser Ile Ile Asn Ile Thr Ala Asn
290 295 300
Leu Pro Ser Thr Gln Asp Leu Leu Ser Phe Leu Gln Ile Gln Leu
305 310 315
Glu Ser Ile Lys Asn Ser Thr Pro Thr Val Val Met Phe Gly Cys
320 325 330
Asp Met Glu Ser Ile Arg Arg Ile Phe Glu Ile Thr Thr Gln Phe
335 340 345
Gly Val Met Pro Pro Glu Leu Arg Trp Val Leu Gly Asp Ser Gln
350 355 360
Asn Val Glu Glu Leu Arg Thr Glu Gly Leu Pro Leu Gly Leu Ile
365 370 375
Ala His Gly Lys Thr Thr Gln Ser Val Phe Glu His Tyr Val Gln
380 385 390
Asp Ala Met Glu Leu Val Ala Arg Ala Val Ala Thr Ala Thr Met
395 400 405
Ile Gln Pro Glu Leu Ala Leu Ile Pro Ser Thr Met Asn Cys Met
410 415 420

```
Glu Val Glu Thr Thr Asn Leu Thr Ser Gly Gln Tyr Leu Ser Arg
                425                 430                 435
Phe Leu Ala Asn Thr Thr Phe Arg Gly Leu Ser Gly Ser Ile Arg
                440                 445                 450
Val Lys Gly Ser Thr Ile Val Ser Ser Glu Asn Asn Phe Phe Ile
                455                 460                 465
Trp Asn Leu Gln His Asp Pro Met Gly Lys Pro Met Trp Thr Arg
                470                 475                 480
Leu Gly Ser Trp Gln Gly Gly Lys Ile Val Met Asp Tyr Gly Ile
                485                 490                 495
Trp Pro Glu Gln Ala Gln Arg His Lys Thr His Phe Gln His Pro
                500                 505                 510
Ser Lys Leu His Leu Arg Val Val Thr Leu Ile Glu His Pro Phe
                515                 520                 525
Val Phe Thr Arg Glu Val Asp Asp Glu Gly Leu Cys Pro Ala Gly
                530                 535                 540
Gln Leu Cys Leu Asp Pro Met Thr Asn Asp Ser Ser Thr Leu Asp
                545                 550                 555
Ser Leu Phe Ser Ser Leu His Ser Ser Asn Asp Thr Val Pro Ile
                560                 565                 570
Lys Phe Lys Lys Cys Cys Tyr Gly Tyr Cys Ile Asp Leu Leu Glu
                575                 580                 585
Lys Ile Ala Glu Asp Met Asn Phe Asp Phe Asp Leu Tyr Ile Val
                590                 595                 600
Gly Asp Gly Lys Tyr Gly Ala Trp Lys Asn Gly His Trp Thr Gly
                605                 610                 615
Leu Val Gly Asp Leu Leu Arg Gly Thr Ala His Met Ala Val Thr
                620                 625                 630
Ser Phe Ser Ile Asn Thr Ala Arg Ser Gln Val Ile Asp Phe Thr
                635                 640                 645
Ser Pro Phe Phe Ser Thr Ser Leu Gly Ile Leu Val Arg Thr Arg
                650                 655                 660
Asp Thr Ala Ala Pro Ile Gly Ala Phe Met Trp Pro Leu His Trp
                665                 670                 675
Thr Met Trp Leu Gly Ile Phe Val Ala Leu His Ile Thr Ala Val
                680                 685                 690
Phe Leu Thr Leu Tyr Glu Trp Lys Ser Pro Phe Gly Leu Thr Ser
                695                 700                 705
Lys Gly Arg Asn Arg Ser Lys Val Phe Ser Phe Ser Ser Ala Leu
                710                 715                 720
Asn Ile Cys Tyr Ala Leu Leu Phe Gly Arg Thr Val Ala Ile Lys
                725                 730                 735
Pro Pro Lys Cys Trp Thr Gly Arg Phe Leu Met Asn Leu Trp Ala
                740                 745                 750
Ile Phe Cys Met Phe Cys Leu Ser Thr Tyr Thr Ala Asn Leu Ala
                755                 760                 765
Ala Val Met Val Gly Glu Lys Ile Tyr Glu Glu Leu Ser Gly Ile
                770                 775                 780
His Asp Pro Lys Leu His His Pro Ser Gln Gly Phe Arg Phe Gly
                785                 790                 795
Thr Val Arg Glu Ser Ser Ala Glu Asp Tyr Val Arg Gln Ser Phe
                800                 805                 810
Pro Glu Met His Glu Tyr Met Arg Arg Tyr Asn Val Pro Ala Thr
                815                 820                 825
Pro Asp Gly Val Glu Tyr Leu Lys Asn Asp Pro Glu Lys Leu Asp
                830                 835                 840
Ala Phe Ile Met Asp Lys Ala Leu Leu Asp Tyr Glu Val Ser Ile
                845                 850                 855
Asp Ala Asp Cys Lys Leu Leu Thr Val Gly Lys Pro Phe Ala Ile
                860                 865                 870
Glu Gly Tyr Gly Ile Gly Leu Pro Pro Asn Ser Pro Leu Thr Ala
                875                 880                 885
Asn Ile Ser Glu Leu Ile Ser Gln Tyr Lys Ser His Gly Phe Met
```

890 895 900
Asp Met Leu His Asp Lys Trp Tyr Arg Val Val Pro Cys Gly Lys
905 910 915
Arg Ser Phe Ala Val Thr Glu Thr Leu Gln Met Gly Ile Lys His
920 925 930
Phe Ser Gly Leu Phe Val Leu Leu Cys Ile Gly Phe Gly Leu Ser
935 940 945
Ile Leu Thr Thr Ile Gly Glu His Ile Val Tyr Arg Leu Leu Leu
950 955 960
Pro Arg Ile Lys Asn Lys Ser Lys Leu Gln Tyr Trp Leu His Thr
965 970 975
Ser Gln Arg Leu His Arg Ala Ile Asn Thr Ser Phe Ile Glu Glu
980 985 990
Lys Gln Gln His Phe Lys Thr Lys Arg Val Glu Lys Arg Ser Asn
995 1000 1005
Val Gly Pro Arg Gln Leu Thr Val Trp Asn Thr Ser Asn Leu Ser
1010 1015 1020
His Asp Asn Arg Arg Lys Tyr Ile Phe Ser Asp Glu Glu Gly Gln
1025 1030 1035
Asn Gln Leu Gly Ile Arg Ile His Gln Asp Ile Pro Leu Pro Pro
1040 1045 1050
Arg Arg Arg Glu Leu Pro Ala Leu Arg Thr Thr Asn Gly Lys Ala
1055 1060 1065
Asp Ser Leu Asn Val Ser Arg Asn Ser Val Met Gln Glu Leu Ser
1070 1075 1080
Glu Leu Glu Lys Gln Ile Gln Val Ile Arg Gln Glu Leu Gln Leu
1085 1090 1095
Ala Val Ser Arg Lys Thr Glu Leu Glu Glu Tyr Gln Arg Thr Ser
1100 1105 1110
Arg Thr Cys Glu Ser
1115

<210> 8
<211> 295
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7472747CD1

<400> 8
Met Pro Ser Ala Gly Leu Cys Ser Cys Trp Gly Gly Arg Val Leu
1 5 10 15
Pro Leu Leu Leu Ala Tyr Val Cys Tyr Leu Leu Leu Gly Ala Thr
20 25 30
Ile Phe Gln Leu Leu Glu Arg Gln Ala Glu Ala Gln Ser Arg Asp
35 40 45
Gln Phe Gln Leu Glu Lys Leu Arg Phe Leu Glu Asn Tyr Thr Cys
50 55 60
Leu Asp Gln Trp Ala Met Glu Gln Phe Val Gln Val Ile Met Glu
65 70 75
Ala Trp Val Lys Gly Val Asn Pro Lys Gly Asn Ser Thr Asn Pro
80 85 90
Ser Asn Trp Asp Phe Gly Ser Ser Phe Phe Phe Ala Gly Thr Val
95 100 105
Val Thr Thr Ile Gly Tyr Gly Asn Leu Ala Pro Ser Thr Glu Ala
110 115 120
Gly Gln Val Phe Cys Val Phe Tyr Ala Leu Leu Gly Ile Pro Leu
125 130 135
Asn Val Ile Phe Leu Asn His Leu Gly Thr Gly Leu Arg Ala His
140 145 150
Leu Ala Ala Ile Glu Arg Trp Glu Asp Arg Pro Arg Arg Ser Gln

```
                155                 160                 165
Glu Val Leu Gln Val Leu Gly Leu Ala Leu Phe Leu Thr Leu Gly
                170                 175                 180
Thr Leu Val Ile Leu Ile Phe Pro Pro Met Val Phe Ser His Val
                185                 190                 195
Glu Gly Trp Ser Phe Ser Glu Gly Phe Tyr Phe Ala Phe Ile Thr
                200                 205                 210
Leu Ser Thr Ile Gly Phe Gly Asp Tyr Val Ala Gly Thr Asp Pro
                215                 220                 225
Ser Lys His Tyr Ile Ser Val Tyr Arg Ser Leu Ala Ala Ile Trp
                230                 235                 240
Ile Leu Leu Gly Leu Ala Trp Leu Ala Leu Ile Leu Pro Leu Gly
                245                 250                 255
Pro Leu Leu Leu His Arg Cys Cys Gln Leu Trp Leu Leu Ser Arg
                260                 265                 270
Gly Leu Gly Val Lys Asp Gly Ala Ala Ser Asp Pro Ser Gly Leu
                275                 280                 285
Pro Arg Pro Gln Lys Ile Pro Ile Ser Ala
                290                 295
```

<210> 9
<211> 384
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7474121CD1

<400> 9

```
Met Glu Val Ser Gly His Pro Gln Ala Arg Arg Cys Cys Pro Glu
 1               5                  10                  15
Ala Leu Gly Lys Leu Phe Pro Gly Leu Cys Phe Leu Cys Phe Leu
                20                  25                  30
Val Thr Tyr Ala Leu Val Gly Ala Val Val Phe Ser Ala Ile Glu
                35                  40                  45
Asp Gly Gln Val Leu Val Ala Ala Asp Asp Gly Glu Phe Glu Lys
                50                  55                  60
Phe Leu Glu Glu Leu Cys Arg Ile Leu Asn Cys Ser Glu Thr Val
                65                  70                  75
Val Glu Asp Arg Lys Gln Asp Leu Gln Gly His Leu Gln Lys Val
                80                  85                  90
Lys Pro Gln Trp Phe Asn Arg Thr Thr His Trp Ser Phe Leu Ser
                95                  100                 105
Ser Leu Phe Phe Cys Cys Thr Val Phe Ser Thr Val Gly Tyr Gly
                110                 115                 120
Tyr Ile Tyr Pro Val Thr Arg Leu Gly Lys Tyr Leu Cys Met Leu
                125                 130                 135
Tyr Ala Leu Phe Gly Ile Pro Leu Met Phe Leu Val Leu Thr Asp
                140                 145                 150
Thr Gly Asp Ile Leu Ala Thr Ile Leu Ser Thr Ser Tyr Asn Arg
                155                 160                 165
Phe Arg Lys Phe Pro Phe Phe Thr Arg Pro Leu Leu Ser Lys Trp
                170                 175                 180
Cys Pro Lys Ser Leu Phe Lys Lys Lys Pro Asp Pro Lys Pro Ala
                185                 190                 195
Asp Glu Ala Val Pro Gln Ile Ile Ile Ser Ala Glu Glu Leu Pro
                200                 205                 210
Gly Pro Lys Leu Gly Thr Cys Pro Ser Arg Pro Ser Cys Ser Met
                215                 220                 225
Glu Leu Phe Glu Arg Ser His Ala Leu Glu Lys Gln Asn Thr Leu
                230                 235                 240
Gln Leu Pro Pro Gln Ala Met Glu Arg Ser Asn Ser Cys Pro Glu
```

245 250 255
Leu Val Leu Gly Arg Leu Ser Tyr Ser Ile Ile Ser Asn Leu Asp
260 265 270
Glu Val Gly Gln Gln Val Glu Arg Leu Asp Ile Pro Leu Pro Ile
275 280 285
Ile Ala Leu Ile Val Phe Ala Tyr Ile Ser Cys Ala Ala Ala Ile
290 295 300
Leu Pro Phe Trp Glu Thr Gln Leu Asp Phe Glu Asn Ala Phe Tyr
305 310 315
Phe Cys Phe Val Thr Leu Thr Thr Ile Gly Phe Gly Asp Thr Val
320 325 330
Leu Glu His Pro Asn Phe Phe Leu Phe Phe Ser Ile Tyr Ile Ile
335 340 345
Val Gly Met Glu Ile Val Phe Ile Ala Phe Lys Leu Val Gln Asn
350 355 360
Arg Leu Ile Asp Ile Tyr Lys Asn Val Met Leu Phe Phe Ala Lys
365 370 375
Gly Lys Phe Tyr His Leu Val Lys Lys
380

<210> 10
<211> 769
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7475615CD1

<400> 10
Met Val Ser Pro Lys Met Tyr Leu Ser Thr Glu Ile Arg Asn Thr
1 5 10 15
Phe Arg Leu Pro Ala Pro Gln Thr His Leu Gln Pro Cys Pro Ala
20 25 30
Gly Phe Ala His Pro Leu Leu Val Asn Ala Pro Asp Met Ser Gln
35 40 45
Pro Arg Pro Arg Tyr Val Val Asp Arg Ala Ala Tyr Ser Leu Thr
50 55 60
Leu Phe Asp Asp Glu Phe Glu Lys Lys Asp Arg Thr Tyr Pro Val
65 70 75
Gly Glu Lys Leu Arg Asn Ala Phe Arg Cys Ser Ser Ala Lys Ile
80 85 90
Lys Ala Val Val Phe Gly Leu Leu Pro Val Leu Ser Trp Leu Pro
95 100 105
Lys Tyr Lys Ile Lys Asp Tyr Ile Ile Pro Asp Leu Leu Gly Gly
110 115 120
Leu Ser Gly Gly Ser Ile Gln Val Pro Gln Gly Met Ala Phe Ala
125 130 135
Leu Leu Ala Asn Leu Pro Ala Val Asn Gly Leu Tyr Ser Ser Phe
140 145 150
Phe Pro Leu Leu Thr Tyr Phe Phe Leu Gly Gly Val His Gln Met
155 160 165
Val Pro Gly Thr Phe Ala Val Ile Ser Ile Leu Val Gly Asn Ile
170 175 180
Cys Leu Gln Leu Ala Pro Glu Ser Lys Phe Gln Val Phe Asn Asn
185 190 195
Ala Thr Asn Glu Ser Tyr Val Asp Thr Ala Ala Met Glu Ala Glu
200 205 210
Arg Leu His Val Ser Ala Thr Leu Ala Cys Leu Thr Ala Ile Ile
215 220 225
Gln Met Gly Leu Gly Phe Met Gln Phe Gly Phe Val Ala Ile Tyr
230 235 240
Leu Ser Glu Ser Phe Ile Arg Gly Phe Met Thr Ala Ala Gly Leu

```
                245                 250                 255
Gln Ile Leu Ile Ser Val Leu Lys Tyr Ile Phe Gly Leu Thr Ile
                260                 265                 270
Pro Ser Tyr Thr Gly Pro Gly Ser Ile Val Phe Thr Phe Ile Asp
                275                 280                 285
Ile Cys Lys Asn Leu Pro His Thr Asn Ile Ala Ser Leu Ile Phe
                290                 295                 300
Ala Leu Ile Ser Gly Ala Phe Leu Val Leu Val Lys Glu Leu Asn
                305                 310                 315
Ala Arg Tyr Met His Lys Ile Arg Phe Pro Ile Pro Thr Glu Met
                320                 325                 330
Ile Val Val Val Val Ala Thr Ala Ile Ser Gly Gly Cys Lys Met
                335                 340                 345
Pro Lys Lys Tyr His Met Gln Ile Val Gly Glu Ile Gln Arg Gly
                350                 355                 360
Phe Pro Thr Pro Val Ser Pro Val Val Ser Gln Trp Lys Asp Met
                365                 370                 375
Ile Gly Thr Ala Phe Ser Leu Ala Ile Val Ser Tyr Val Ile Asn
                380                 385                 390
Leu Ala Met Gly Arg Thr Leu Ala Asn Lys His Gly Tyr Asp Val
                395                 400                 405
Asp Ser Asn Gln Glu Met Ile Ala Leu Gly Cys Ser Asn Phe Phe
                410                 415                 420
Gly Ser Phe Phe Lys Ile His Val Ile Cys Cys Ala Leu Ser Val
                425                 430                 435
Thr Leu Ala Val Asp Gly Ala Gly Gly Lys Ser Gln Ser Val Leu
                440                 445                 450
Gly Ala Leu Ile Ala Val Asn Leu Lys Asn Ser Leu Lys Gln Leu
                455                 460                 465
Thr Asp Pro Tyr Tyr Leu Trp Arg Lys Ser Lys Leu Asp Cys Cys
                470                 475                 480
Ile Trp Val Val Ser Phe Leu Ser Ser Phe Phe Leu Ser Leu Pro
                485                 490                 495
Tyr Gly Val Ala Val Gly Val Ala Phe Ser Val Leu Val Val Val
                500                 505                 510
Phe Gln Thr Gln Phe Arg Asn Gly Tyr Ala Leu Ala Gln Val Met
                515                 520                 525
Asp Thr Asp Ile Tyr Val Asn Pro Lys Thr Tyr Asn Arg Ala Gln
                530                 535                 540
Asp Ile Gln Gly Ile Lys Ile Ile Thr Tyr Cys Ser Pro Leu Tyr
                545                 550                 555
Phe Ala Asn Ser Glu Ile Phe Arg Gln Lys Val Ile Ala Lys Thr
                560                 565                 570
Val Ser Leu Gln Glu Leu Gln Gln Asp Phe Glu Asn Ala Pro Pro
                575                 580                 585
Thr Asp Pro Asn Asn Asn Gln Thr Pro Ala Asn Gly Thr Ser Val
                590                 595                 600
Ser Tyr Ile Thr Phe Ser Pro Asp Ser Ser Ser Pro Ala Gln Ser
                605                 610                 615
Glu Pro Pro Ala Ser Ala Glu Ala Pro Gly Glu Pro Ser Asp Met
                620                 625                 630
Leu Ala Ser Val Pro Pro Phe Val Thr Phe His Thr Leu Ile Leu
                635                 640                 645
Asp Met Ser Gly Val Ser Phe Val Asp Leu Met Gly Ile Lys Ala
                650                 655                 660
Leu Ala Lys Leu Ser Ser Thr Tyr Gly Lys Ile Gly Val Lys Val
                665                 670                 675
Phe Leu Val Asn Ile His Ala Gln Val Tyr Asn Asp Ile Ser His
                680                 685                 690
Gly Gly Val Phe Glu Asp Gly Ser Leu Glu Cys Lys His Val Phe
                695                 700                 705
Pro Ser Ile His Asp Ala Val Leu Phe Ala Gln Ala Asn Ala Arg
                710                 715                 720
```

Asp Val Thr Pro Gly His Asn Phe Gln Gly Ala Pro Gly Asp Ala
725 730 735
Glu Leu Ser Leu Tyr Asp Ser Glu Glu Asp Ile Arg Ser Tyr Trp
740 745 750
Asp Leu Glu Gln Glu Met Phe Gly Ser Met Phe His Ala Glu Thr
755 760 765
Leu Thr Ala Leu

<210> 11
<211> 882
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7475656CD1

<400> 11
Met Glu Gly Gly Gly Lys Pro Asn Ser Ser Ser Asn Ser Arg Asp
1 5 10 15
Asp Gly Asn Ser Val Phe Pro Ala Lys Ala Ser Ala Pro Gly Ala
20 25 30
Gly Pro Ala Ala Ala Glu Lys Arg Leu Gly Thr Pro Pro Gly Gly
35 40 45
Gly Gly Ala Gly Ala Lys Glu His Gly Asn Ser Val Cys Phe Lys
50 55 60
Val Asp Gly Gly Gly Gly Glu Glu Pro Ala Gly Gly Phe Glu Asp
65 70 75
Ala Glu Gly Pro Arg Arg Gln Tyr Gly Phe Met Gln Arg Gln Phe
80 85 90
Thr Ser Met Leu Gln Pro Gly Val Asn Lys Phe Ser Leu Arg Met
95 100 105
Phe Gly Ser Gln Lys Ala Val Glu Lys Glu Gln Glu Arg Val Lys
110 115 120
Thr Ala Gly Phe Trp Ile Ile His Pro Tyr Ser Asp Phe Arg Phe
125 130 135
Tyr Trp Asp Leu Ile Met Leu Ile Met Met Val Gly Asn Leu Val
140 145 150
Ile Ile Pro Val Gly Ile Thr Phe Phe Thr Glu Gln Thr Thr Thr
155 160 165
Pro Trp Ile Ile Phe Asn Val Ala Ser Asp Thr Val Phe Leu Leu
170 175 180
Asp Leu Ile Met Asn Phe Arg Thr Gly Thr Val Asn Glu Asp Ser
185 190 195
Ser Glu Ile Ile Leu Asp Pro Lys Val Ile Lys Met Asn Tyr Leu
200 205 210
Lys Ser Trp Phe Val Val Asp Phe Ile Ser Ser Ile Pro Val Asp
215 220 225
Tyr Ile Phe Leu Ile Val Glu Lys Gly Met Asp Ser Glu Val Tyr
230 235 240
Lys Thr Ala Arg Ala Leu Arg Ile Val Arg Phe Thr Lys Ile Leu
245 250 255
Ser Leu Leu Arg Leu Leu Arg Leu Ser Arg Leu Ile Arg Tyr Ile
260 265 270
His Gln Trp Glu Glu Ile Phe His Met Thr Tyr Asp Leu Ala Ser
275 280 285
Ala Val Val Arg Ile Phe Asn Leu Ile Gly Met Met Leu Leu Leu
290 295 300
Cys His Trp Asp Gly Cys Leu Gln Phe Leu Val Pro Leu Leu Gln
305 310 315
Asp Phe Pro Pro Asp Cys Trp Val Ser Leu Asn Glu Met Val Asn
320 325 330

```
Asp Ser Trp Gly Lys Gln Tyr Ser Tyr Ala Leu Phe Lys Ala Met
                335                 340                 345
Ser His Met Leu Cys Ile Gly Tyr Gly Ala Gln Ala Pro Val Ser
                350                 355                 360
Met Ser Asp Leu Trp Ile Thr Met Leu Ser Met Ile Val Gly Ala
                365                 370                 375
Thr Cys Tyr Ala Met Phe Val Gly His Ala Thr Ala Leu Ile Gln
                380                 385                 390
Ser Leu Asp Ser Ser Arg Arg Gln Tyr Gln Glu Lys Tyr Lys Gln
                395                 400                 405
Val Glu Gln Tyr Met Ser Phe His Lys Leu Pro Ala Asp Met Arg
                410                 415                 420
Gln Lys Ile His Asp Tyr Tyr Glu His Arg Tyr Gln Gly Lys Ile
                425                 430                 435
Phe Asp Glu Glu Asn Ile Leu Asn Glu Leu Asn Asp Pro Leu Arg
                440                 445                 450
Glu Glu Ile Val Asn Phe Asn Cys Arg Lys Leu Val Ala Thr Met
                455                 460                 465
Pro Leu Phe Ala Asn Ala Asp Pro Asn Phe Val Thr Ala Met Leu
                470                 475                 480
Ser Lys Leu Arg Phe Glu Val Phe Gln Pro Gly Asp Tyr Ile Ile
                485                 490                 495
Arg Glu Gly Ala Val Gly Lys Lys Met Tyr Phe Ile Gln His Gly
                500                 505                 510
Val Ala Gly Val Ile Thr Lys Ser Ser Lys Glu Met Lys Leu Thr
                515                 520                 525
Asp Gly Ser Tyr Phe Gly Glu Ile Cys Leu Leu Thr Lys Gly Arg
                530                 535                 540
Arg Thr Ala Ser Val Arg Ala Asp Thr Tyr Cys Arg Leu Tyr Ser
                545                 550                 555
Leu Ser Val Asp Asn Phe Asn Glu Val Leu Glu Glu Tyr Pro Met
                560                 565                 570
Met Arg Arg Ala Phe Glu Thr Val Ala Ile Asp Arg Leu Asp Arg
                575                 580                 585
Ile Gly Lys Lys Asn Ser Ile Leu Leu Gln Lys Phe Gln Lys Asp
                590                 595                 600
Leu Asn Thr Gly Val Phe Asn Asn Gln Glu Asn Glu Ile Leu Lys
                605                 610                 615
Gln Ile Val Lys His Asp Arg Glu Met Val Gln Ala Ile Ala Pro
                620                 625                 630
Ile Asn Tyr Pro Gln Met Thr Thr Leu Asn Ser Thr Ser Ser Thr
                635                 640                 645
Thr Thr Pro Thr Ser Arg Met Arg Thr Gln Ser Pro Pro Val Tyr
                650                 655                 660
Thr Ala Thr Ser Leu Ser His Ser Asn Leu His Ser Pro Ser Pro
                665                 670                 675
Ser Thr Gln Thr Pro Gln Pro Ser Ala Ile Leu Ser Pro Cys Ser
                680                 685                 690
Tyr Thr Thr Ala Val Cys Ser Pro Pro Val Gln Ser Pro Leu Ala
                695                 700                 705
Ala Arg Thr Phe His Tyr Ala Ser Pro Thr Ala Ser Gln Leu Ser
                710                 715                 720
Leu Met Gln Gln Gln Pro Gln Gln Gln Val Gln Gln Ser Gln Pro
                725                 730                 735
Pro Gln Thr Gln Pro Gln Gln Pro Ser Pro Gln Pro Gln Thr Pro
                740                 745                 750
Gly Ser Ser Thr Pro Lys Asn Glu Val His Lys Ser Thr Gln Ala
                755                 760                 765
Leu His Asn Thr Asn Leu Thr Arg Glu Val Arg Pro Leu Ser Ala
                770                 775                 780
Ser Gln Pro Ser Leu Pro His Glu Val Ser Thr Leu Ile Ser Arg
                785                 790                 795
Pro His Pro Thr Val Gly Glu Ser Leu Ala Ser Ile Pro Gln Pro
```

```
                    800                 805                 810
Val Thr Ala Val Pro Gly Thr Gly Leu Gln Ala Gly Gly Arg Ser
                    815                 820                 825
Thr Val Pro Gln Arg Val Thr Leu Phe Arg Gln Met Ser Ser Gly
                    830                 835                 840
Ala Ile Pro Pro Asn Arg Gly Val Pro Pro Ala Pro Pro Pro Pro
                    845                 850                 855
Ala Ala Ala Leu Pro Arg Glu Ser Ser Ser Val Leu Asn Thr Asp
                    860                 865                 870
Pro Asp Ala Glu Lys Pro Arg Phe Ala Ser Asn Leu
                    875                 880

<210> 12
<211> 1547
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7480632CD1

<400> 12
Met Val Lys Lys Glu Ile Ser Val Arg Gln Gln Ile Gln Ala Leu
  1               5                  10                  15
Leu Tyr Lys Asn Phe Leu Lys Lys Trp Arg Ile Lys Arg Glu Phe
                 20                  25                  30
Leu Glu Glu Trp Thr Ile Thr Leu Phe Leu Gly Leu Tyr Leu Cys
                 35                  40                  45
Ile Phe Ser Glu His Phe Arg Ala Thr Arg Phe Pro Glu Gln Pro
                 50                  55                  60
Pro Lys Val Leu Gly Ser Val Asp Gln Phe Asn Asp Ser Gly Leu
                 65                  70                  75
Val Val Ala Tyr Thr Pro Val Ser Asn Ile Thr Gln Arg Ile Met
                 80                  85                  90
Asn Lys Met Ala Leu Ala Ser Phe Met Lys Gly Arg Thr Val Ile
                 95                 100                 105
Gly Thr Pro Asp Glu Glu Thr Met Asp Ile Glu Leu Pro Lys Lys
                110                 115                 120
Tyr His Glu Met Val Gly Val Ile Phe Ser Asp Thr Phe Ser Tyr
                125                 130                 135
Arg Leu Lys Phe Asn Trp Gly Tyr Arg Ile Pro Val Ile Lys Glu
                140                 145                 150
His Ser Glu Tyr Thr Gly His Cys Trp Ala Met His Gly Glu Ile
                155                 160                 165
Phe Cys Tyr Leu Ala Lys Tyr Trp Leu Lys Gly Phe Val Ala Phe
                170                 175                 180
Gln Ala Ala Ile Asn Ala Ala Ile Ile Glu Val Thr Thr Asn His
                185                 190                 195
Ser Val Met Glu Glu Leu Thr Ser Val Ile Gly Ile Asn Met Lys
                200                 205                 210
Ile Pro Pro Phe Ile Ser Lys Gly Glu Ile Met Asn Glu Trp Phe
                215                 220                 225
His Phe Thr Cys Leu Val Ser Phe Ser Ser Phe Ile Tyr Phe Ala
                230                 235                 240
Ser Leu Asn Val Ala Arg Glu Arg Gly Lys Phe Lys Lys Leu Met
                245                 250                 255
Thr Val Met Gly Leu Arg Glu Ser Ala Phe Trp Leu Ser Trp Gly
                260                 265                 270
Leu Thr Tyr Ile Cys Phe Ile Phe Ile Met Ser Ile Phe Met Ala
                275                 280                 285
Leu Val Ile Thr Ser Ile Pro Ile Val Phe His Thr Gly Phe Met
                290                 295                 300
Val Ile Phe Thr Leu Tyr Ser Leu Tyr Gly Leu Ser Leu Val Ala
```

```
                305                     310                     315
Leu Ala Phe Leu Met Ser Val Leu Ile Arg Lys Pro Met Leu Ala
                320                     325                     330
Gly Leu Ala Gly Phe Leu Phe Thr Val Phe Trp Gly Cys Leu Gly
                335                     340                     345
Phe Thr Val Leu Tyr Arg Gln Leu Pro Leu Ser Leu Gly Trp Val
                350                     355                     360
Leu Ser Leu Leu Ser Pro Phe Ala Phe Thr Ala Gly Met Ala Gln
                365                     370                     375
Ile Thr His Leu Asp Asn Tyr Leu Ser Gly Val Ile Phe Pro Asp
                380                     385                     390
Pro Ser Gly Asp Ser Tyr Lys Met Ile Ala Thr Phe Phe Ile Leu
                395                     400                     405
Ala Phe Asp Thr Leu Phe Tyr Leu Ile Phe Thr Leu Tyr Phe Glu
                410                     415                     420
Arg Val Leu Pro Gly Lys Asp Gly His Gly Asp Ser Pro Leu Phe
                425                     430                     435
Phe Leu Lys Ser Ser Phe Trp Ser Lys His Gln Asn Thr His His
                440                     445                     450
Glu Ile Phe Glu Asn Glu Ile Asn Pro Glu His Ser Ser Asp Asp
                455                     460                     465
Ser Phe Glu Pro Val Ser Pro Glu Phe His Gly Lys Glu Ala Ile
                470                     475                     480
Arg Ile Arg Asn Val Ile Lys Glu Tyr Asn Gly Lys Thr Gly Lys
                485                     490                     495
Val Glu Ala Leu Gln Gly Ile Phe Phe Asp Ile Tyr Glu Gly Gln
                500                     505                     510
Ile Thr Ala Ile Leu Gly His Asn Gly Ala Gly Lys Ser Thr Leu
                515                     520                     525
Leu Asn Ile Leu Ser Gly Leu Ser Val Ser Thr Glu Gly Ser Ala
                530                     535                     540
Thr Ile Tyr Asn Thr Gln Leu Ser Glu Ile Thr Asp Met Glu Glu
                545                     550                     555
Ile Arg Lys Asn Ile Gly Phe Cys Pro Gln Phe Asn Phe Gln Phe
                560                     565                     570
Asp Phe Leu Thr Val Arg Glu Asn Leu Arg Val Phe Ala Lys Ile
                575                     580                     585
Lys Gly Ile Gln Pro Lys Glu Val Glu Gln Glu Val Leu Leu Leu
                590                     595                     600
Asp Glu Pro Thr Ala Gly Leu Asp Pro Phe Ser Arg His Arg Val
                605                     610                     615
Trp Ser Leu Leu Lys Glu His Lys Val Asp Arg Leu Ile Leu Phe
                620                     625                     630
Ser Thr Gln Phe Met Asp Glu Ala Asp Ile Leu Ala Asp Arg Lys
                635                     640                     645
Val Phe Leu Ser Asn Gly Lys Leu Lys Cys Ala Gly Ser Ser Leu
                650                     655                     660
Phe Leu Lys Arg Lys Trp Gly Ile Gly Tyr His Leu Ser Leu His
                665                     670                     675
Arg Asn Glu Met Cys Asp Thr Glu Lys Ile Thr Ser Leu Ile Lys
                680                     685                     690
Gln His Ile Pro Asp Ala Lys Leu Thr Thr Glu Ser Glu Glu Lys
                695                     700                     705
Leu Val Tyr Ser Leu Pro Leu Glu Lys Thr Asn Lys Phe Pro Asp
                710                     715                     720
Leu Tyr Ser Asp Leu Asp Lys Cys Ser Asp Gln Gly Ile Arg Asn
                725                     730                     735
Tyr Ala Val Ser Val Thr Ser Leu Asn Glu Val Phe Leu Asn Leu
                740                     745                     750
Glu Gly Lys Ser Ala Ile Asp Glu Pro Asp Phe Asp Ile Gly Lys
                755                     760                     765
Gln Glu Lys Ile His Val Thr Arg Asn Thr Gly Asp Glu Ser Glu
                770                     775                     780
```

Met Glu Gln Val Leu Cys Ser Leu Pro Glu Thr Arg Lys Ala Val
785 790 795
Ser Ser Ala Ala Leu Trp Arg Arg Gln Ile Tyr Ala Val Ala Thr
800 805 810
Leu Arg Phe Leu Lys Leu Arg Arg Glu Arg Arg Ala Leu Leu Cys
815 820 825
Leu Leu Leu Val Leu Gly Ile Ala Phe Ile Pro Ile Ile Leu Glu
830 835 840
Lys Ile Met Tyr Lys Val Thr Arg Glu Thr His Cys Trp Glu Phe
845 850 855
Ser Pro Ser Met Tyr Phe Leu Ser Leu Glu Gln Ile Pro Lys Thr
860 865 870
Pro Leu Thr Ser Leu Leu Ile Val Asn Asn Thr Gly Ser Asn Ile
875 880 885
Glu Asp Leu Val His Ser Leu Lys Cys Gln Asp Ile Val Leu Glu
890 895 900
Ile Asp Asp Phe Arg Asn Arg Asn Gly Ser Asp Asp Pro Ser Tyr
905 910 915
Asn Gly Ala Ile Ile Val Ser Gly Asp Gln Lys Asp Tyr Arg Phe
920 925 930
Ser Val Ala Cys Asn Thr Lys Lys Leu Asn Cys Phe Pro Val Leu
935 940 945
Met Gly Ile Val Ser Asn Ala Leu Met Gly Ile Phe Asn Phe Thr
950 955 960
Glu Leu Ile Gln Met Glu Ser Thr Ser Phe Phe Phe Tyr Ile Thr
965 970 975
Thr Lys Ser Phe Gln Thr Lys Ile Pro Ser Ser Ile Pro Ser Ile
980 985 990
Leu Cys Gln Lys Asn Val Gln Ser Gln Leu Trp Ile Ser Gly Leu
995 1000 1005
Trp Pro Ser Ala Tyr Trp Cys Gly Gln Ala Leu Val Asp Ile Pro
1010 1015 1020
Leu Tyr Phe Leu Ile Leu Phe Ser Ile His Leu Ile Tyr Tyr Phe
1025 1030 1035
Ile Phe Leu Gly Phe Gln Leu Ser Trp Glu Leu Met Phe Val Leu
1040 1045 1050
Val Val Cys Ile Ile Gly Cys Ala Val Ser Leu Ile Phe Leu Thr
1055 1060 1065
Tyr Val Leu Ser Phe Ile Phe Arg Lys Trp Arg Lys Asn Asn Gly
1070 1075 1080
Phe Trp Ser Phe Gly Phe Phe Ile Val Ser Ile Tyr Thr Asp Phe
1085 1090 1095
Ser Phe His Tyr Asn Val Ser Arg Cys Asp Phe Leu Phe Ile Phe
1100 1105 1110
Ile Phe Val Cys Leu Phe Ile Ala His His Phe Ser Phe Cys Ser
1115 1120 1125
Pro Tyr Leu Gln Ser Val Ile Phe Leu Phe Val Ile Arg Cys Leu
1130 1135 1140
Glu Met Lys Tyr Gly Asn Glu Ile Met Asn Lys Asp Pro Val Phe
1145 1150 1155
Arg Ile Ser Pro Arg Ser Arg Glu Thr His Pro Asn Pro Glu Glu
1160 1165 1170
Pro Glu Glu Glu Asp Glu Asp Val Gln Ala Glu Arg Val Gln Ala
1175 1180 1185
Ala Asn Ala Leu Thr Ala Pro Asn Leu Glu Glu Glu Pro Val Ile
1190 1195 1200
Thr Ala Ser Cys Leu His Lys Glu Tyr Tyr Glu Thr Lys Lys Ser
1205 1210 1215
Cys Phe Ser Thr Arg Lys Lys Lys Ile Ala Ile Arg Asn Val Ser
1220 1225 1230
Phe Cys Val Lys Lys Gly Glu Val Leu Gly Leu Leu Gly His Asn
1235 1240 1245
Gly Ala Gly Lys Ser Thr Ser Ile Lys Met Ile Thr Gly Cys Thr

```
                1250                    1255                    1260
Lys Pro Thr Ala Gly Val Val Val Leu Gln Gly Ser Arg Ala Ser
                1265                    1270                    1275
Val Arg Gln Gln His Asp Asn Ser Leu Lys Phe Leu Gly Tyr Cys
                1280                    1285                    1290
Pro Gln Glu Asn Ser Leu Trp Pro Lys Leu Thr Met Lys Glu His
                1295                    1300                    1305
Leu Glu Leu Tyr Ala Ala Val Lys Gly Leu Gly Lys Glu Asp Ala
                1310                    1315                    1320
Ala Leu Ser Ile Ser Arg Leu Val Glu Ala Leu Lys Leu Gln Glu
                1325                    1330                    1335
Gln Leu Lys Ala Pro Val Lys Thr Leu Ser Glu Gly Ile Lys Arg
                1340                    1345                    1350
Lys Leu Cys Phe Val Leu Ser Ile Leu Gly Asn Pro Ser Val Val
                1355                    1360                    1365
Leu Leu Asp Glu Pro Phe Thr Gly Met Asp Pro Glu Gly Gln Gln
                1370                    1375                    1380
Gln Met Trp Gln Ile Leu Gln Ala Thr Val Lys Asn Lys Glu Arg
                1385                    1390                    1395
Gly Thr Leu Leu Thr Thr His Tyr Met Ser Glu Ala Glu Ala Val
                1400                    1405                    1410
Cys Asp Arg Met Ala Met Met Val Ser Gly Thr Leu Arg Cys Ile
                1415                    1420                    1425
Gly Ser Ile Gln His Leu Lys Asn Lys Phe Gly Arg Asp Tyr Leu
                1430                    1435                    1440
Leu Glu Ile Lys Met Lys Glu Pro Thr Gln Val Glu Ala Leu His
                1445                    1450                    1455
Thr Glu Ile Leu Lys Leu Phe Pro Gln Ala Ala Trp Gln Glu Arg
                1460                    1465                    1470
Tyr Ser Ser Leu Met Ala Tyr Lys Leu Pro Val Glu Asp Val His
                1475                    1480                    1485
Pro Leu Ser Arg Ala Phe Phe Lys Leu Glu Ala Met Lys Gln Thr
                1490                    1495                    1500
Phe Asn Leu Glu Glu Tyr Ser Leu Ser Gln Ala Thr Leu Glu Gln
                1505                    1510                    1515
Val Phe Leu Glu Leu Cys Lys Glu Gln Glu Leu Gly Asn Val Asp
                1520                    1525                    1530
Asp Lys Ile Asp Thr Thr Val Glu Trp Lys Leu Leu Pro Gln Glu
                1535                    1540                    1545
Asp Pro


<210> 13
<211> 698
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 6952742CD1

<400> 13
Met Asp Glu Ser Pro Glu Pro Leu Gln Gln Gly Arg Gly Pro Val
 1               5                  10                  15
Pro Val Arg Arg Gln Arg Pro Ala Pro Arg Gly Leu Arg Glu Met
                20                  25                  30
Leu Lys Ala Arg Leu Trp Cys Ser Cys Ser Cys Ser Val Leu Cys
                35                  40                  45
Val Arg Ala Leu Val Gln Asp Leu Leu Pro Ala Thr Arg Trp Leu
                50                  55                  60
Arg Gln Tyr Arg Pro Arg Glu Tyr Leu Ala Gly Asp Val Met Ser
                65                  70                  75
Gly Leu Val Ile Gly Ile Ile Leu Ala Ile Ala Tyr Ser Leu Leu
```

```
                80                  85                  90
Ala Gly Leu Gln Pro Ile Tyr Ser Leu Tyr Thr Ser Phe Phe Ala
                95                  100                 105
Asn Leu Ile Tyr Phe Leu Met Gly Thr Ser Arg His Val Ser Val
                110                 115                 120
Gly Ile Phe Ser Leu Leu Cys Leu Met Val Gly Gln Val Val Asp
                125                 130                 135
Arg Glu Leu Gln Leu Ala Gly Phe Asp Pro Ser Gln Asp Gly Leu
                140                 145                 150
Gln Pro Gly Ala Asn Ser Ser Thr Leu Asn Gly Ser Ala Ala Met
                155                 160                 165
Leu Asp Cys Gly Arg Asp Cys Tyr Ala Ile Arg Val Ala Thr Ala
                170                 175                 180
Leu Thr Leu Met Thr Gly Leu Tyr Gln Val Leu Met Gly Val Leu
                185                 190                 195
Arg Leu Gly Phe Val Ser Ala Tyr Leu Ser Gln Pro Leu Leu Asp
                200                 205                 210
Gly Phe Ala Met Gly Ala Ser Val Thr Ile Leu Thr Ser Gln Leu
                215                 220                 225
Lys His Leu Leu Gly Val Arg Ile Pro Arg His Gln Gly Pro Gly
                230                 235                 240
Met Val Val Leu Thr Trp Leu Ser Leu Leu Arg Gly Ala Gly Gln
                245                 250                 255
Ala Asn Val Cys Asp Val Val Thr Ser Thr Val Cys Leu Ala Val
                260                 265                 270
Leu Leu Ala Ala Lys Glu Leu Ser Asp Arg Tyr Arg His Arg Leu
                275                 280                 285
Arg Val Pro Leu Pro Thr Glu Leu Leu Val Ile Val Val Ala Thr
                290                 295                 300
Leu Val Ser His Phe Gly Gln Leu His Lys Arg Phe Gly Ser Ser
                305                 310                 315
Val Ala Gly Asp Ile Pro Thr Gly Phe Met Pro Pro Gln Val Pro
                320                 325                 330
Glu Pro Arg Leu Met Gln Arg Val Ala Leu Asp Ala Val Ala Leu
                335                 340                 345
Ala Leu Val Ala Ala Ala Phe Ser Ile Ser Leu Ala Glu Met Phe
                350                 355                 360
Ala Arg Ser His Gly Tyr Ser Val Arg Ala Asn Gln Glu Leu Leu
                365                 370                 375
Ala Val Gly Cys Cys Asn Val Leu Pro Ala Phe Leu His Cys Phe
                380                 385                 390
Ala Thr Ser Ala Ala Leu Ala Lys Ser Leu Val Lys Thr Ala Thr
                395                 400                 405
Gly Cys Arg Thr Gln Leu Ser Ser Val Val Ser Ala Thr Val Val
                410                 415                 420
Leu Leu Val Leu Leu Ala Leu Ala Pro Leu Phe His Asp Leu Gln
                425                 430                 435
Arg Ser Val Leu Ala Cys Val Ile Val Val Ser Leu Arg Gly Ala
                440                 445                 450
Leu Arg Lys Val Trp Asp Leu Pro Arg Leu Trp Arg Met Ser Pro
                455                 460                 465
Ala Asp Ala Leu Val Trp Ala Gly Thr Val Ala Thr Cys Met Leu
                470                 475                 480
Val Ser Thr Glu Ala Gly Leu Leu Ala Gly Val Ile Leu Ser Leu
                485                 490                 495
Leu Ser Leu Ala Gly Arg Thr Gln Ser His Gly Thr Ala Leu Leu
                500                 505                 510
Ala Arg Ile Gly Asp Thr Ala Phe Tyr Glu Asp Ala Thr Glu Phe
                515                 520                 525
Glu Gly Leu Val Pro Glu Pro Gly Val Arg Val Phe Arg Phe Gly
                530                 535                 540
Gly Pro Leu Tyr Tyr Ala Asn Lys Asp Phe Phe Leu Gln Ser Leu
                545                 550                 555
```

```
Tyr Ser Leu Thr Gly Leu Asp Ala Gly Cys Met Ala Ala Arg Arg
                560                 565                 570
Lys Glu Gly Gly Ser Glu Thr Gly Val Gly Glu Gly Gly Pro Ala
                575                 580                 585
Gln Gly Glu Asp Leu Gly Pro Val Ser Thr Arg Ala Ala Leu Val
                590                 595                 600
Pro Ala Ala Ala Gly Phe His Thr Val Val Ile Asp Cys Ala Pro
                605                 610                 615
Leu Leu Phe Leu Asp Ala Ala Gly Val Ser Thr Leu Gln Asp Leu
                620                 625                 630
Arg Arg Asp Tyr Gly Ala Leu Gly Ile Ser Leu Leu Leu Ala Cys
                635                 640                 645
Cys Ser Pro Pro Val Arg Asp Ile Leu Ser Arg Gly Gly Phe Leu
                650                 655                 660
Gly Glu Gly Pro Gly Asp Thr Ala Glu Glu Glu Gln Leu Phe Leu
                665                 670                 675
Ser Val His Asp Ala Val Gln Thr Ala Arg Ala Arg His Arg Glu
                680                 685                 690
Leu Glu Ala Thr Asp Ala His Leu
                695

<210> 14
<211> 766
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7478795CD1

<400> 14
Met Arg Leu Trp Lys Ala Val Val Val Thr Leu Ala Phe Met Ser
 1               5                  10                  15
Val Asp Ile Cys Val Thr Thr Ala Ile Tyr Val Phe Ser His Leu
                20                  25                  30
Asp Arg Ser Leu Leu Glu Asp Ile Arg His Phe Asn Ile Phe Asp
                35                  40                  45
Ser Val Leu Asp Leu Trp Ala Ala Cys Leu Tyr Arg Ser Cys Leu
                50                  55                  60
Leu Leu Gly Ala Thr Ile Gly Val Ala Lys Asn Ser Ala Leu Gly
                65                  70                  75
Pro Arg Arg Leu Arg Ala Ser Trp Leu Val Ile Thr Leu Val Cys
                80                  85                . 90
Leu Phe Val Gly Ile Tyr Ala Met Val Lys Leu Leu Leu Phe Ser
                95                  100                 105
Glu Val Arg Arg Pro Ile Arg Asp Pro Trp Phe Trp Ala Leu Phe
                110                 115                 120
Val Trp Thr Tyr Ile Ser Leu Gly Ala Ser Phe Leu Leu Trp Trp
                125                 130                 135
Leu Leu Ser Thr Val Arg Pro Gly Thr Gln Ala Leu Glu Pro Gly
                140                 145                 150
Ala Ala Thr Glu Ala Glu Gly Phe Pro Gly Ser Gly Arg Pro Pro
                155                 160                 165
Pro Glu Gln Ala Ser Gly Ala Thr Leu Gln Lys Leu Leu Ser Tyr
                170                 175                 180
Thr Lys Pro Asp Val Ala Phe Leu Val Ala Ala Ser Phe Phe Leu
                185                 190                 195
Ile Val Ala Ala Leu Gly Glu Thr Phe Leu Pro Tyr Tyr Thr Gly
                200                 205                 210
Arg Ala Ile Asp Gly Ile Val Ile Gln Lys Ser Met Asp Gln Phe
                215                 220                 225
Ser Thr Ala Val Val Ile Val Cys Leu Leu Ala Ile Gly Ser Ser
                230                 235                 240
```

Phe Ala Ala Gly Ile Arg Gly Gly Ile Phe Thr Leu Ile Phe Ala
245 250 255
Arg Leu Asn Ile Arg Leu Arg Asn Cys Leu Phe Arg Ser Leu Val
260 265 270
Ser Gln Glu Thr Ser Phe Phe Asp Glu Asn Arg Thr Gly Asp Leu
275 280 285
Ile Ser Arg Leu Thr Ser Asp Thr Thr Met Val Ser Asp Leu Val
290 295 300
Ser Gln Asn Ile Asn Val Phe Leu Arg Asn Thr Val Lys Val Thr
305 310 315
Gly Val Val Val Phe Met Phe Ser Leu Ser Trp Gln Leu Ser Leu
320 325 330
Val Thr Phe Met Gly Phe Pro Ile Ile Met Met Val Ser Asn Ile
335 340 345
Tyr Gly Lys Tyr Tyr Lys Arg Leu Ser Lys Glu Val Gln Asn Ala
350 355 360
Leu Ala Arg Ala Ser Asn Thr Ala Glu Glu Thr Ile Ser Ala Met
365 370 375
Lys Thr Val Arg Ser Phe Ala Asn Glu Glu Glu Glu Ala Glu Val
380 385 390
Tyr Leu Arg Lys Leu Gln Gln Val Tyr Lys Leu Asn Arg Lys Glu
395 400 405
Ala Ala Ala Tyr Met Tyr Tyr Val Trp Gly Ser Gly Leu Thr Leu
410 415 420
Leu Val Val Gln Val Ser Ile Leu Tyr Tyr Gly Gly His Leu Val
425 430 435
Ile Ser Gly Gln Met Thr Ser Gly Asn Leu Ile Ala Phe Ile Ile
440 445 450
Tyr Glu Phe Val Leu Gly Asp Cys Met Glu Ser Val Gly Ser Val
455 460 465
Tyr Ser Gly Leu Met Gln Gly Val Gly Ala Ala Glu Lys Val Phe
470 475 480
Glu Phe Ile Asp Arg Gln Pro Thr Met Val His Asp Gly Ser Leu
485 490 495
Ala Pro Asp His Leu Glu Gly Arg Val Asp Phe Glu Asn Val Thr
500 505 510
Phe Thr Tyr Arg Thr Arg Pro His Thr Gln Val Leu Gln Asn Val
515 520 525
Ser Phe Ser Leu Ser Pro Gly Lys Val Thr Ala Leu Val Gly Pro
530 535 540
Ser Gly Ser Gly Lys Ser Ser Cys Val Asn Ile Leu Glu Asn Phe
545 550 555
Tyr Pro Leu Glu Gly Gly Arg Val Leu Leu Asp Gly Lys Pro Ile
560 565 570
Ser Ala Tyr Asp His Lys Tyr Leu His Arg Val Ile Ser Leu Val
575 580 585
Ser Gln Glu Pro Val Leu Phe Ala Arg Ser Ile Thr Asp Asn Ile
590 595 600
Ser Tyr Gly Leu Pro Thr Val Pro Phe Glu Met Val Val Glu Ala
605 610 615
Ala Gln Lys Ala Asn Ala His Gly Phe Ile Met Glu Leu Gln Asp
620 625 630
Gly Tyr Ser Thr Glu Thr Gly Glu Lys Gly Ala Gln Leu Ser Gly
635 640 645
Gly Gln Lys Gln Arg Val Ala Met Ala Arg Ala Leu Val Arg Asn
650 655 660
Pro Pro Val Leu Ile Leu Asp Glu Ala Thr Ser Ala Leu Asp Ala
665 670 675
Glu Ser Glu Tyr Leu Ile Gln Gln Ala Ile His Gly Asn Leu Gln
680 685 690
Lys His Thr Val Leu Ile Ile Ala His Arg Leu Ser Thr Val Glu
695 700 705
His Ala His Leu Ile Val Val Leu Asp Lys Gly Arg Val Val Gln

```
                710                     715                     720
Gln Gly Thr His Gln Gln Leu Leu Ala Gln Gly Gly Leu Tyr Ala
                725                     730                     735
Lys Leu Val Gln Arg Gln Met Leu Gly Leu Gln Pro Ala Ala Asp
                740                     745                     750
Phe Thr Ala Gly His Asn Glu Pro Val Ala Asn Gly Ser His Lys
                755                     760                     765
Ala


&lt;210&gt; 15
&lt;211&gt; 450
&lt;212&gt; PRT
&lt;213&gt; Homo sapiens

&lt;220&gt;
&lt;221&gt; misc_feature
&lt;223&gt; Incyte ID No: 656293CD1

&lt;400&gt; 15
Met Gly Leu Arg Ser His His Leu Ser Leu Gly Leu Leu Leu Leu
 1               5                   10                  15
Phe Leu Leu Pro Ala Glu Cys Leu Gly Ala Glu Gly Arg Leu Ala
                20                  25                  30
Leu Lys Leu Phe Arg Asp Leu Phe Ala Asn Tyr Thr Ser Ala Leu
                35                  40                  45
Arg Pro Val Ala Asp Thr Asp Gln Thr Leu Asn Val Thr Leu Glu
                50                  55                  60
Val Thr Leu Ser Gln Ile Ile Asp Met Asp Glu Arg Asn Gln Val
                65                  70                  75
Leu Thr Leu Tyr Leu Trp Ile Arg Gln Glu Trp Thr Asp Ala Tyr
                80                  85                  90
Leu Arg Trp Asp Pro Asn Ala Tyr Gly Gly Leu Asp Ala Ile Arg
                95                  100                 105
Ile Pro Ser Ser Leu Val Trp Arg Pro Asp Ile Val Leu Tyr Asn
                110                 115                 120
Lys Ala Asp Ala Gln Pro Pro Gly Ser Ala Ser Thr Asn Val Val
                125                 130                 135
Leu Arg His Asp Gly Ala Val Arg Trp Asp Ala Pro Ala Ile Thr
                140                 145                 150
Arg Ser Ser Cys Arg Val Asp Val Ala Ala Phe Pro Phe Asp Ala
                155                 160                 165
Gln His Cys Gly Leu Thr Phe Gly Ser Trp Thr His Gly Gly His
                170                 175                 180
Gln Leu Asp Val Arg Pro Arg Gly Ala Ala Ala Ser Leu Ala Asp
                185                 190                 195
Phe Val Glu Asn Val Glu Trp Arg Val Leu Gly Met Pro Ala Arg
                200                 205                 210
Arg Arg Val Leu Thr Tyr Gly Cys Cys Ser Glu Pro Tyr Pro Asp
                215                 220                 225
Val Thr Phe Thr Leu Leu Leu Arg Arg Arg Ala Ala Ala Tyr Val
                230                 235                 240
Cys Asn Leu Leu Leu Pro Cys Val Leu Ile Ser Leu Leu Ala Pro
                245                 250                 255
Leu Ala Phe His Leu Pro Ala Asp Ser Gly Glu Lys Val Ser Leu
                260                 265                 270
Gly Val Thr Val Leu Leu Ala Leu Thr Val Phe Gln Leu Leu Leu
                275                 280                 285
Ala Glu Ser Met Pro Pro Ala Glu Ser Val Pro Leu Ile Gly Lys
                290                 295                 300
Tyr Tyr Met Ala Thr Met Thr Met Val Thr Phe Ser Thr Ala Leu
                305                 310                 315
Thr Ile Leu Ile Met Asn Leu His Tyr Cys Gly Pro Ser Val Arg
```

```
                320                 325                 330
Pro Val Pro Ala Trp Ala Arg Ala Leu Leu Leu Gly His Leu Ala
                335                 340                 345
Arg Gly Leu Cys Val Arg Glu Arg Gly Glu Pro Cys Gly Gln Ser
                350                 355                 360
Arg Pro Pro Glu Leu Ser Pro Ser Pro Gln Ser Pro Glu Gly Gly
                365                 370                 375
Ala Gly Pro Pro Ala Gly Pro Cys His Glu Pro Arg Cys Leu Cys
                380                 385                 390
Arg Gln Glu Ala Leu Leu His His Val Ala Thr Ile Ala Asn Thr
                395                 400                 405
Phe Arg Ser His Arg Ala Ala Gln Arg Cys His Glu Asp Trp Lys
                410                 415                 420
Arg Leu Ala Arg Val Met Asp Arg Phe Phe Leu Ala Ile Phe Phe
                425                 430                 435
Ser Met Ala Leu Val Met Ser Leu Leu Val Leu Val Gln Ala Leu
                440                 445                 450
```

<210> 16
<211> 260
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7473957CD1

<400> 16

```
Met Pro Ile Leu Ala Asn Leu Pro Gly Met Ser Ser Pro Arg Ala
 1               5                  10                  15
Met Glu Phe Thr Ser Ser Gly Ser Ala Asn Thr Glu Thr Thr Lys
                20                  25                  30
Val Thr Gly Ser Leu Glu Thr Lys Tyr Arg Trp Thr Glu Tyr Gly
                35                  40                  45
Leu Thr Phe Thr Glu Lys Trp Asn Thr Asp Asn Thr Leu Gly Thr
                50                  55                  60
Glu Ile Thr Val Glu Asp Gln Leu Ala Arg Gly Leu Lys Leu Thr
                65                  70                  75
Phe Asp Ser Ser Phe Ser Pro Asn Thr Gly Lys Lys Asn Ala Lys
                80                  85                  90
Ile Lys Thr Gly Tyr Lys Arg Glu His Ile Asn Leu Gly Cys Asp
                95                 100                 105
Met Asp Phe Asp Ile Ala Gly Pro Ser Ile Arg Gly Ala Leu Val
                110                 115                 120
Leu Gly Tyr Glu Gly Trp Leu Ala Gly Tyr Gln Met Asn Phe Glu
                125                 130                 135
Thr Ala Lys Ser Arg Val Thr Gln Ser Asn Phe Ala Val Gly Tyr
                140                 145                 150
Lys Thr Asp Glu Phe Gln Leu His Thr Asn Val Asn Asp Gly Thr
                155                 160                 165
Glu Phe Gly Gly Ser Ile Tyr Gln Lys Val Asn Lys Lys Leu Glu
                170                 175                 180
Thr Ala Val Asn Leu Ala Trp Thr Ala Gly Asn Ser Asn Thr Arg
                185                 190                 195
Phe Gly Ile Ala Ala Lys Tyr Gln Ile Asp Pro Asp Ala Cys Phe
                200                 205                 210
Ser Ala Lys Val Asn Asn Ser Ser Leu Ile Gly Leu Gly Tyr Thr
                215                 220                 225
Gln Thr Leu Lys Pro Gly Ile Lys Leu Thr Leu Ser Ala Leu Leu
                230                 235                 240
Asp Gly Lys Asn Val Asn Ala Gly Gly His Lys Leu Gly Leu Gly
                245                 250                 255
```

Leu Glu Phe Gln Ala
260

<210> 17
<211> 506
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7474111CD1

<400> 17
Met Ser Glu Pro Glu Leu Gly Ser Gly Gln Phe Leu Glu Lys Ala
1 5 10 15
Leu Gln Thr Pro Ser Val Pro Ala Pro Glu Ser Thr Leu Gly Phe
20 25 30
Glu Pro Gly Leu Leu Lys Gly Ala Leu Gly Thr Ala Gln Phe Ile
35 40 45
Pro Met Ala Gln Gly Arg Thr Arg Glu Gln Ala Ser Arg Arg Trp
50 55 60
Ala Pro Arg Ser Pro Ala Leu Arg Thr Pro Pro Arg His Tyr Gly
65 70 75
Pro Glu Arg Arg Gly Arg Thr Ala Ser Arg Gly Gly Glu Pro Glu
80 85 90
Val Gln Gly Gly Ala Pro Gly Asn Pro Ser Pro Ser Lys Pro Gly
95 100 105
Ser Pro Gln Gly Val Gly Pro Ala Ala Trp Glu Arg Ala Pro Arg
110 115 120
Pro Arg Cys Ala Gln Pro Ser Gly Ala Arg Val Gly Glu Arg Thr
125 130 135
Gln Pro Arg Ser Gln Pro Val Gly Leu Ser Arg Gly Ala Gly Glu
140 145 150
Asp Ser Pro Ala Thr Arg Ser Gly Ala Ala Ser Val Val Leu Asn
155 160 165
Val Gly Gly Ala Arg Tyr Ser Leu Ser Arg Glu Leu Leu Lys Asp
170 175 180
Phe Pro Leu Arg Arg Val Ser Arg Leu His Gly Cys Arg Ser Glu
185 190 195
Arg Asp Val Leu Glu Val Cys Asp Asp Tyr Asp Arg Glu Arg Asn
200 205 210
Glu Tyr Phe Phe Asp Arg His Ser Glu Ala Phe Gly Phe Ile Leu
215 220 225
Leu Tyr Ala Ala Pro Ser Arg Arg Trp Leu Glu Arg Met Arg Arg
230 235 240
Thr Phe Glu Glu Pro Thr Ser Ser Leu Ala Ala Gln Ile Leu Ala
245 250 255
Ser Val Ser Val Val Phe Val Ile Val Ser Met Val Val Leu Cys
260 265 270
Ala Ser Thr Leu Pro Asp Trp Arg Asn Ala Ala Ala Asp Asn Arg
275 280 285
Ser Leu Asp Asp Arg Ser Arg Ile Ile Glu Ala Ile Cys Ile Gly
290 295 300
Trp Phe Thr Ala Glu Cys Ile Val Arg Phe Ile Val Ser Lys Asn
305 310 315
Lys Cys Glu Phe Val Lys Arg Pro Leu Asn Ile Ile Asp Leu Leu
320 325 330
Ala Ile Thr Pro Tyr Tyr Ile Ser Val Leu Met Thr Val Phe Thr
335 340 345
Gly Glu Asn Ser Gln Leu Gln Arg Ala Gly Val Thr Leu Arg Val
350 355 360
Leu Arg Met Met Arg Ile Phe Trp Val Ile Lys Leu Ala Arg His
365 370 375

```
Phe Ile Gly Leu Gln Thr Leu Gly Leu Thr Leu Lys Arg Cys Tyr
                380                 385                 390
Arg Glu Met Val Met Leu Leu Val Phe Ile Cys Val Ala Met Ala
                395                 400                 405
Ile Phe Ser Ala Leu Ser Gln Leu Leu Glu His Gly Leu Asp Leu
                410                 415                 420
Glu Thr Ser Asn Lys Asp Phe Thr Ser Ile Pro Ala Ala Cys Trp
                425                 430                 435
Trp Val Ile Ile Ser Met Thr Thr Val Gly Tyr Gly Asp Met Tyr
                440                 445                 450
Pro Ile Thr Val Pro Gly Arg Ile Leu Gly Gly Val Cys Val Val
                455                 460                 465
Ser Gly Ile Val Leu Leu Ala Leu Pro Ile Thr Phe Ile Tyr His
                470                 475                 480
Ser Phe Val Gln Cys Tyr His Glu Leu Lys Phe Arg Ser Ala Arg
                485                 490                 495
Tyr Ser Arg Ser Leu Ser Thr Glu Phe Leu Asn
                500                 505

<210> 18
<211> 506
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7480826CD1

<400> 18
Met Lys Lys Ala Glu Met Gly Arg Phe Ser Ile Ser Pro Asp Glu
  1               5                  10                  15
Asp Ser Ser Ser Tyr Ser Ser Asn Ser Asp Phe Asn Tyr Ser Tyr
                 20                  25                  30
Pro Thr Lys Gln Ala Ala Leu Lys Ser His Tyr Ala Asp Val Asp
                 35                  40                  45
Pro Glu Asn Gln Asn Phe Leu Leu Glu Ser Asn Leu Gly Lys Lys
                 50                  55                  60
Lys Tyr Glu Thr Glu Phe His Pro Gly Thr Thr Ser Phe Gly Met
                 65                  70                  75
Ser Val Phe Asn Leu Ser Asn Ala Ile Val Gly Ser Gly Ile Leu
                 80                  85                  90
Gly Leu Ser Tyr Ala Met Ala Asn Thr Gly Ile Ala Leu Phe Ile
                 95                 100                 105
Ile Leu Leu Thr Phe Val Ser Ile Phe Ser Leu Tyr Ser Val His
                110                 115                 120
Leu Leu Leu Lys Thr Ala Asn Glu Gly Gly Ser Leu Leu Tyr Glu
                125                 130                 135
Gln Leu Gly Tyr Lys Ala Phe Gly Leu Val Gly Lys Leu Ala Ala
                140                 145                 150
Ser Gly Ser Ile Thr Met Gln Asn Ile Gly Ala Met Ser Ser Tyr
                155                 160                 165
Leu Phe Ile Val Lys Tyr Glu Leu Pro Leu Val Ile Gln Ala Leu
                170                 175                 180
Thr Asn Ile Glu Asp Lys Thr Gly Leu Trp Tyr Leu Asn Gly Asn
                185                 190                 195
Tyr Leu Val Leu Leu Val Ser Leu Val Val Ile Leu Pro Leu Ser
                200                 205                 210
Leu Phe Arg Asn Leu Gly Tyr Leu Gly Tyr Thr Ser Gly Leu Ser
                215                 220                 225
Leu Leu Cys Met Val Phe Phe Leu Ile Val Val Ile Cys Lys Lys
                230                 235                 240
Phe Gln Val Pro Cys Pro Val Glu Ala Ala Leu Ile Ile Asn Glu
                245                 250                 255
```

Thr Ile Asn Thr Thr Leu Thr Gln Pro Thr Ala Leu Val Pro Ala
260 265 270
Leu Ser His Asn Val Thr Glu Asn Asp Ser Cys Arg Pro His Tyr
275 280 285
Phe Ile Phe Asn Ser Gln Thr Val Tyr Ala Val Pro Ile Leu Ile
290 295 300
Phe Ser Phe Val Cys His Pro Ala Val Leu Pro Ile Tyr Glu Glu
305 310 315
Leu Lys Asp Arg Ser Arg Arg Arg Met Met Asn Val Ser Lys Ile
320 325 330
Ser Phe Phe Ala Met Phe Leu Met Tyr Leu Leu Ala Ala Leu Phe
335 340 345
Gly Tyr Leu Thr Phe Tyr Glu His Val Glu Ser Glu Leu Leu His
350 355 360
Thr Tyr Ser Ser Ile Leu Gly Thr Asp Ile Leu Leu Leu Ile Val
365 370 375
Arg Leu Ala Val Leu Met Ala Val Thr Leu Thr Val Pro Val Val
380 385 390
Ile Phe Pro Ile Arg Ser Ser Val Thr His Leu Leu Cys Ala Ser
395 400 405
Lys Asp Phe Ser Trp Trp Arg His Ser Leu Ile Thr Val Ser Ile
410 415 420
Leu Ala Phe Thr Asn Leu Leu Val Ile Phe Val Pro Thr Ile Arg
425 430 435
Asp Ile Phe Gly Phe Ile Gly Ala Ser Ala Ala Ser Met Leu Ile
440 445 450
Phe Ile Leu Pro Ser Ala Phe Tyr Ile Lys Leu Val Lys Lys Glu
455 460 465
Pro Met Lys Ser Val Gln Lys Ile Gly Ala Leu Phe Phe Leu Leu
470 475 480
Ser Gly Val Leu Val Met Thr Gly Ser Met Ala Leu Ile Val Leu
485 490 495
Asp Trp Val His Asn Ala Pro Gly Gly Gly His
500 505

<210> 19
<211> 315
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 6025572CD1

<400> 19
Met His Arg Glu Pro Ala Lys Lys Lys Ala Glu Lys Arg Leu Phe
1 5 10 15
Asp Ala Ser Ser Phe Gly Lys Asp Leu Leu Ala Gly Gly Val Ala
20 25 30
Ala Ala Val Ser Lys Thr Ala Val Ala Pro Ile Glu Arg Val Lys
35 40 45
Leu Leu Leu Gln Val Gln Ala Ser Ser Lys Gln Ile Ser Pro Glu
50 55 60
Ala Arg Tyr Lys Gly Met Val Asp Cys Leu Val Arg Ile Pro Arg
65 70 75
Glu Gln Gly Phe Phe Ser Phe Trp Arg Gly Asn Leu Ala Asn Val
80 85 90
Ile Arg Tyr Phe Pro Thr Gln Ala Leu Asn Phe Ala Phe Lys Asp
95 100 105
Lys Tyr Lys Gln Leu Phe Met Ser Gly Val Asn Lys Glu Lys Gln
110 115 120
Phe Trp Arg Trp Phe Leu Ala Asn Leu Ala Ser Gly Gly Ala Ala
125 130 135

Gly Ala Thr Ser Leu Cys Val Val Tyr Pro Leu Asp Phe Ala Arg
140 145 150
Thr Arg Leu Gly Val Asp Ile Gly Lys Gly Pro Glu Glu Arg Gln
155 160 165
Phe Lys Gly Leu Gly Asp Cys Ile Met Lys Ile Ala Lys Ser Asp
170 175 180
Gly Ile Ala Gly Leu Tyr Gln Gly Phe Gly Val Ser Val Gln Gly
185 190 195
Ile Ile Val Tyr Arg Ala Ser Tyr Phe Gly Ala Tyr Asp Thr Val
200 205 210
Lys Gly Leu Leu Pro Lys Pro Lys Lys Thr Pro Phe Leu Val Ser
215 220 225
Phe Phe Ile Ala Gln Val Val Thr Thr Cys Ser Gly Ile Leu Ser
230 235 240
Tyr Pro Phe Asp Thr Val Arg Arg Arg Met Met Met Gln Ser Gly
245 250 255
Glu Ala Lys Arg Gln Tyr Lys Gly Thr Leu Asp Cys Phe Val Lys
260 265 270
Ile Tyr Gln His Glu Gly Ile Ser Ser Phe Phe Arg Gly Ala Phe
275 280 285
Ser Asn Val Leu Arg Gly Thr Gly Gly Ala Leu Val Leu Val Leu
290 295 300
Tyr Asp Lys Ile Lys Glu Phe Phe His Ile Asp Ile Gly Gly Arg
305 310 315

<210> 20
<211> 540
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 5686561CD1

<400> 20
Met Val Pro Ala Gly Trp Val Arg Gly Leu Glu Leu Ser Leu Trp
1 5 10 15
Gly Gly Asp Pro Val Val Pro Trp Ser Cys Arg Phe Cys Ser Gln
20 25 30
Gln Asp Asp Gly Gln Asp Arg Glu Arg Leu Thr Tyr Phe Gln Asn
35 40 45
Leu Pro Glu Ser Leu Thr Ser Leu Leu Val Leu Leu Thr Thr Ala
50 55 60
Asn Asn Pro Asp Val Met Ile Pro Ala Tyr Ser Lys Asn Arg Ala
65 70 75
Tyr Ala Ile Phe Phe Ile Val Phe Thr Val Ile Gly Ser Leu Phe
80 85 90
Leu Met Asn Leu Leu Thr Ala Ile Ile Tyr Ser Gln Phe Arg Gly
95 100 105
Tyr Leu Met Lys Ser Leu Gln Thr Ser Leu Phe Arg Arg Arg Leu
110 115 120
Gly Thr Arg Ala Ala Phe Glu Val Leu Ser Ser Met Val Gly Glu
125 130 135
Gly Gly Ala Phe Pro Gln Ala Val Gly Val Lys Pro Gln Asn Leu
140 145 150
Leu Gln Val Leu Gln Lys Val Gln Leu Asp Ser Ser His Lys Gln
155 160 165
Ala Met Met Glu Lys Val Arg Ser Tyr Asp Ser Val Leu Leu Ser
170 175 180
Ala Glu Glu Phe Gln Lys Leu Phe Asn Glu Leu Asp Arg Ser Val
185 190 195
Val Lys Glu His Pro Pro Arg Pro Glu Tyr Gln Ser Pro Phe Leu

200 205 210
Gln Ser Ala Gln Phe Leu Phe Gly His Tyr Tyr Phe Asp Tyr Leu
215 220 225
Gly Asn Leu Ile Ala Leu Ala Asn Leu Val Ser Ile Cys Val Phe
230 235 240
Leu Val Leu Asp Ala Asp Val Leu Pro Ala Glu Arg Asp Asp Phe
245 250 255
Ile Leu Gly Ile Leu Asn Cys Val Phe Ile Val Tyr Tyr Leu Leu
260 265 270
Glu Met Leu Leu Lys Val Phe Ala Leu Gly Leu Arg Gly Tyr Leu
275 280 285
Ser Tyr Pro Ser Asn Val Phe Asp Gly Leu Leu Thr Val Val Leu
290 295 300
Leu Val Leu Glu Ile Ser Thr Leu Ala Val Tyr Arg Leu Pro His
305 310 315
Pro Gly Trp Arg Pro Glu Met Val Gly Leu Leu Ser Leu Trp Asp
320 325 330
Met Thr Arg Met Leu Asn Met Leu Ile Val Phe Arg Phe Leu Arg
335 340 345
Ile Ile Pro Ser Met Lys Pro Met Ala Val Val Ala Ser Thr Val
350 355 360
Leu Gly Leu Val Gln Asn Met Arg Ala Phe Gly Gly Ile Leu Val
365 370 375
Val Val Tyr Tyr Val Phe Ala Ile Ile Gly Ile Asn Leu Phe Arg
380 385 390
Gly Val Ile Val Ala Leu Pro Gly Asn Ser Ser Leu Ala Pro Ala
395 400 405
Asn Gly Ser Ala Pro Cys Gly Ser Phe Glu Gln Leu Glu Tyr Trp
410 415 420
Ala Asn Asn Phe Asp Asp Phe Ala Ala Ala Leu Val Thr Leu Trp
425 430 435
Asn Leu Met Val Val Asn Asn Trp Gln Val Phe Leu Asp Ala Tyr
440 445 450
Arg Arg Tyr Ser Gly Pro Trp Ser Lys Ile Tyr Phe Val Leu Trp
455 460 465
Trp Leu Val Ser Ser Val Ile Trp Val Asn Leu Phe Leu Ala Leu
470 475 480
Ile Leu Glu Asn Phe Leu His Lys Trp Asp Pro Arg Ser His Leu
485 490 495
Gln Pro Leu Ala Gly Thr Pro Glu Ala Thr Tyr Gln Met Thr Val
500 505 510
Glu Leu Leu Phe Arg Asp Ile Leu Glu Glu Pro Gly Glu Asp Glu
515 520 525
Leu Thr Glu Arg Leu Ser Gln His Pro His Leu Trp Leu Cys Arg
530 535 540

<210> 21
<211> 322
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1553725CD1

<400> 21
Met Glu Ala Asp Leu Ser Gly Phe Asn Ile Asp Ala Pro Arg Trp
1 5 10 15
Asp Gln Arg Thr Phe Leu Gly Arg Val Lys His Phe Leu Asn Ile
20 25 30
Thr Asp Pro Arg Thr Val Phe Val Ser Glu Arg Glu Leu Asp Trp
35 40 45

```
Ala Lys Val Met Val Glu Lys Ser Arg Met Gly Val Val Pro Pro
                50                  55                  60
Gly Thr Gln Val Glu Gln Leu Leu Tyr Ala Lys Lys Leu Tyr Asp
                65                  70                  75
Ser Ala Phe His Pro Asp Thr Gly Glu Lys Met Asn Val Ile Gly
                80                  85                  90
Arg Met Ser Phe Gln Leu Pro Gly Gly Met Ile Ile Thr Gly Phe
                95                  100                 105
Met Leu Gln Phe Tyr Arg Thr Met Pro Ala Val Ile Phe Trp Gln
                110                 115                 120
Trp Val Asn Gln Ser Phe Asn Ala Leu Val Asn Tyr Thr Asn Arg
                125                 130                 135
Asn Ala Ala Ser Pro Thr Ser Val Arg Gln Met Ala Leu Ser Tyr
                140                 145                 150
Phe Thr Ala Thr Thr Thr Ala Val Ala Thr Ala Val Gly Met Asn
                155                 160                 165
Met Leu Thr Lys Lys Ala Pro Pro Leu Val Gly Arg Trp Val Pro
                170                 175                 180
Phe Ala Ala Val Ala Ala Ala Asn Cys Val Asn Ile Pro Met Met
                185                 190                 195
Arg Gln Gln Glu Leu Ile Lys Gly Ile Cys Val Lys Asp Arg Asn
                200                 205                 210
Glu Asn Glu Ile Gly His Ser Arg Arg Ala Ala Ala Ile Gly Ile
                215                 220                 225
Thr Gln Val Val Ile Ser Arg Ile Thr Met Ser Ala Pro Gly Met
                230                 235                 240
Ile Leu Leu Pro Val Ile Met Glu Arg Leu Glu Lys Leu His Phe
                245                 250                 255
Met Gln Lys Val Lys Val Leu His Ala Pro Leu Gln Val Met Leu
                260                 265                 270
Ser Gly Cys Phe Leu Ile Phe Met Val Pro Val Ala Cys Gly Leu
                275                 280                 285
Phe Pro Gln Lys Cys Glu Leu Pro Val Ser Tyr Leu Glu Pro Lys
                290                 295                 300
Leu Gln Asp Thr Ile Lys Ala Lys Tyr Gly Glu Leu Glu Pro Tyr
                305                 310                 315
Val Tyr Phe Asn Lys Gly Leu
                320
```

```
<210> 22
<211> 417
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1695770CD1

<400> 22
Met Thr Thr Leu Val Pro Ala Thr Leu Ser Phe Leu Leu Leu Trp
 1               5                  10                  15
Thr Leu Pro Gly Gln Val Leu Leu Arg Val Ala Leu Ala Lys Glu
                20                  25                  30
Glu Val Lys Ser Gly Thr Lys Gly Ser Gln Pro Met Ser Pro Ser
                35                  40                  45
Asp Phe Leu Asp Lys Leu Met Gly Arg Thr Ser Gly Tyr Asp Ala
                50                  55                  60
Arg Ile Arg Pro Asn Phe Lys Gly Pro Pro Val Asn Val Thr Cys
                65                  70                  75
Asn Ile Phe Ile Asn Ser Phe Ser Ser Val Thr Lys Thr Thr Met
                80                  85                  90
Asp Tyr Arg Val Asn Val Phe Leu Arg Gln Gln Trp Asn Asp Pro
                95                  100                 105
```

Arg Leu Ser Tyr Arg Glu Tyr Pro Asp Asp Ser Leu Asp Leu Asp
110 115 120
Pro Ser Met Leu Asp Ser Ile Trp Lys Pro Asp Leu Phe Phe Ala
125 130 135
Asn Glu Lys Gly Ala Asn Phe His Glu Val Thr Thr Asp Asn Lys
140 145 150
Leu Leu Arg Ile Phe Lys Asn Gly Asn Val Leu Tyr Ser Ile Arg
155 160 165
Leu Thr Leu Ile Leu Ser Cys Leu Met Asp Leu Lys Asn Phe Pro
170 175 180
Met Asp Ile Gln Thr Cys Thr Met Gln Leu Glu Ser Phe Gly Tyr
185 190 195
Thr Met Lys Asp Leu Val Phe Glu Trp Leu Glu Asp Ala Pro Ala
200 205 210
Val Gln Val Ala Glu Gly Leu Thr Leu Pro Gln Phe Ile Leu Arg
215 220 225
Asp Glu Lys Asp Leu Gly Cys Cys Thr Lys His Tyr Asn Thr Gly
230 235 240
Lys Phe Thr Cys Ile Glu Val Lys Phe His Leu Glu Arg Gln Met
245 250 255
Gly Tyr Tyr Leu Ile Gln Met Tyr Ile Pro Ser Leu Leu Ile Val
260 265 270
Ile Leu Ser Trp Val Ser Phe Trp Ile Asn Met Asp Ala Ala Pro
275 280 285
Ala Arg Val Gly Leu Gly Ile Thr Thr Val Leu Thr Met Thr Thr
290 295 300
Gln Ser Ser Gly Ser Arg Ala Ser Leu Pro Lys Val Ser Tyr Val
305 310 315
Lys Ala Ile Asp Ile Trp Met Ala Val Cys Leu Leu Phe Val Phe
320 325 330
Ala Ala Leu Leu Glu Tyr Ala Ala Ile Asn Phe Val Ser Arg Gln
335 340 345
His Lys Glu Phe Ile Arg Leu Arg Arg Arg Gln Arg Arg Gln Arg
350 355 360
Leu Glu Glu Asp Ile Ile Gln Glu Ser Arg Phe Tyr Phe Arg Gly
365 370 375
Tyr Gly Leu Gly His Cys Leu Gln Ala Arg Asp Gly Gly Pro Met
380 385 390
Glu Gly Ser Gly Ile Tyr Ser Pro Gln Pro Pro Ala Pro Leu Leu
395 400 405
Arg Glu Gly Glu Thr Thr Arg Lys Leu Tyr Val Asp
410 415

<210> 23
<211> 1864
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 4672222CD1

<400> 23
Met Ser Gln Lys Ser Trp Ile Glu Ser Thr Leu Thr Lys Arg Glu
1 5 10 15
Cys Val Tyr Ile Ile Pro Ser Ser Lys Asp Pro His Arg Cys Leu
20 25 30
Pro Gly Cys Gln Ile Cys Gln Gln Leu Val Arg Cys Phe Cys Gly
35 40 45
Arg Leu Val Lys Gln His Ala Cys Phe Thr Ala Ser Leu Ala Met
50 55 60
Lys Tyr Ser Asp Val Lys Leu Gly Asp His Phe Asn Gln Ala Ile
65 70 75

Glu Glu Trp Ser Val Glu Lys His Thr Glu Gln Ser Pro Thr Asp
80 85 90
Ala Tyr Gly Val Ile Asn Phe Gln Gly Gly Ser His Ser Tyr Arg
95 100 105
Ala Lys Tyr Val Arg Leu Ser Tyr Asp Thr Lys Pro Glu Val Ile
110 115 120
Leu Gln Leu Leu Leu Lys Glu Trp Gln Met Glu Leu Pro Lys Leu
125 130 135
Val Ile Ser Val His Gly Gly Met Gln Lys Phe Glu Leu His Pro
140 145 150
Arg Ile Lys Gln Leu Leu Gly Lys Gly Leu Ile Lys Ala Ala Val
155 160 165
Thr Thr Gly Ala Trp Ile Leu Thr Gly Gly Val Asn Thr Gly Val
170 175 180
Ala Lys His Val Gly Asp Ala Leu Lys Glu His Ala Ser Arg Ser
185 190 195
Ser Arg Lys Ile Cys Thr Ile Gly Ile Ala Pro Trp Gly Val Ile
200 205 210
Glu Asn Arg Asn Asp Leu Val Gly Arg Asp Val Val Ala Pro Tyr
215 220 225
Gln Thr Leu Leu Asn Pro Leu Ser Lys Leu Asn Val Leu Asn Asn
230 235 240
Leu His Ser His Phe Ile Leu Val Asp Asp Gly Thr Val Gly Lys
245 250 255
Tyr Gly Ala Glu Val Arg Leu Arg Arg Glu Leu Glu Lys Thr Ile
260 265 270
Asn Gln Gln Arg Ile His Ala Arg Ile Gly Gln Gly Val Pro Val
275 280 285
Val Ala Leu Ile Phe Glu Gly Gly Pro Asn Val Ile Leu Thr Val
290 295 300
Leu Glu Tyr Leu Gln Glu Ser Pro Pro Val Pro Val Val Val Cys
305 310 315
Glu Gly Thr Gly Arg Ala Ala Asp Leu Leu Ala Tyr Ile His Lys
320 325 330
Gln Thr Glu Glu Gly Gly Asn Leu Pro Asp Ala Ala Glu Pro Asp
335 340 345
Ile Ile Ser Thr Ile Lys Lys Thr Phe Asn Phe Gly Gln Asn Glu
350 355 360
Ala Leu His Leu Phe Gln Thr Leu Met Glu Cys Met Lys Arg Lys
365 370 375
Glu Leu Ile Thr Val Phe His Ile Gly Ser Asp Glu His Gln Asp
380 385 390
Ile Asp Val Ala Ile Leu Thr Ala Leu Leu Lys Gly Thr Asn Ala
395 400 405
Ser Ala Phe Asp Gln Leu Ile Leu Thr Leu Ala Trp Asp Arg Val
410 415 420
Asp Ile Ala Lys Asn His Val Phe Val Tyr Gly Gln Gln Trp Leu
425 430 435
Val Gly Ser Leu Glu Gln Ala Met Leu Asp Ala Leu Val Met Asp
440 445 450
Arg Val Ala Phe Val Lys Leu Leu Ile Glu Asn Gly Val Ser Met
455 460 465
His Lys Phe Leu Thr Ile Pro Arg Leu Glu Glu Leu Tyr Asn Thr
470 475 480
Lys Gln Gly Pro Thr Asn Pro Met Leu Phe His Leu Val Arg Asp
485 490 495
Val Lys Gln Gly Asn Leu Pro Pro Gly Tyr Lys Ile Thr Leu Ile
500 505 510
Asp Ile Gly Leu Val Ile Glu Tyr Leu Met Gly Gly Thr Tyr Arg
515 520 525
Cys Thr Tyr Thr Arg Lys Arg Phe Arg Leu Ile Tyr Asn Ser Leu
530 535 540
Gly Gly Asn Asn Arg Arg Ser Gly Arg Asn Thr Ser Ser Ser Thr

545 550 555
Pro Gln Leu Arg Lys Ser His Glu Ser Phe Gly Asn Arg Ala Asp
560 565 570
Lys Lys Glu Lys Met Arg His Asn His Phe Ile Lys Thr Ala Gln
575 580 585
Pro Tyr Arg Pro Lys Ile Asp Thr Val Met Glu Glu Gly Lys Lys
590 595 600
Lys Arg Thr Lys Asp Glu Ile Val Asp Ile Asp Asp Pro Glu Thr
605 610 615
Lys Arg Phe Pro Tyr Pro Leu Asn Glu Leu Leu Ile Trp Ala Cys
620 625 630
Leu Met Lys Arg Gln Val Met Ala Arg Phe Leu Trp Gln His Gly
635 640 645
Glu Glu Ser Met Ala Lys Ala Leu Val Ala Cys Lys Ile Tyr Arg
650 655 660
Ser Met Ala Tyr Glu Ala Lys Gln Ser Asp Leu Val Asp Asp Thr
665 670 675
Ser Glu Glu Leu Lys Gln Tyr Ser Asn Asp Phe Gly Gln Leu Ala
680 685 690
Val Glu Leu Leu Glu Gln Ser Phe Arg Gln Asp Glu Thr Met Ala
695 700 705
Met Lys Leu Leu Thr Tyr Glu Leu Lys Asn Trp Ser Asn Ser Thr
710 715 720
Cys Leu Lys Leu Ala Val Ser Ser Arg Leu Arg Pro Phe Val Ala
725 730 735
His Thr Cys Thr Gln Met Leu Leu Ser Asp Met Trp Met Gly Arg
740 745 750
Leu Asn Met Arg Lys Asn Ser Trp Tyr Lys Val Ile Leu Ser Ile
755 760 765
Leu Val Pro Pro Ala Ile Leu Leu Leu Glu Tyr Lys Thr Lys Ala
770 775 780
Glu Met Ser His Ile Pro Gln Ser Gln Asp Ala His Gln Met Thr
785 790 795
Met Asp Asp Ser Glu Asn Asn Phe Gln Asn Ile Thr Glu Glu Ile
800 805 810
Pro Met Glu Val Phe Lys Glu Val Arg Ile Leu Asp Ser Asn Glu
815 820 825
Gly Lys Asn Glu Met Glu Ile Gln Met Lys Ser Lys Lys Leu Pro
830 835 840
Ile Thr Arg Lys Phe Tyr Ala Phe Tyr His Ala Pro Ile Val Lys
845 850 855
Phe Trp Phe Asn Thr Leu Ala Tyr Leu Gly Phe Leu Met Leu Tyr
860 865 870
Thr Phe Val Val Leu Val Gln Met Glu Gln Leu Pro Ser Val Gln
875 880 885
Glu Trp Ile Val Ile Ala Tyr Ile Phe Thr Tyr Ala Ile Glu Lys
890 895 900
Val Arg Glu Ile Phe Met Ser Glu Ala Gly Lys Val Asn Gln Lys
905 910 915
Ile Lys Val Trp Phe Ser Asp Tyr Phe Asn Ile Ser Asp Thr Ile
920 925 930
Ala Ile Ile Ser Phe Phe Ile Gly Phe Gly Leu Arg Phe Gly Ala
935 940 945
Lys Trp Asn Phe Ala Asn Ala Tyr Asp Asn His Val Phe Val Ala
950 955 960
Gly Arg Leu Ile Tyr Cys Leu Asn Ile Ile Phe Trp Tyr Val Arg
965 970 975
Leu Leu Asp Phe Leu Ala Val Asn Gln Gln Ala Gly Pro Tyr Val
980 985 990
Met Met Ile Gly Lys Met Val Ala Asn Met Phe Tyr Ile Val Val
995 1000 1005
Ile Met Ala Leu Val Leu Leu Ser Phe Gly Val Pro Arg Lys Ala
1010 1015 1020

Ile Leu Tyr Pro His Glu Ala Pro Ser Trp Thr Leu Ala Lys Asp
1025 1030 1035
Ile Val Phe His Pro Tyr Trp Met Ile Phe Gly Glu Val Tyr Ala
1040 1045 1050
Tyr Glu Ile Asp Val Cys Ala Asn Asp Ser Val Ile Pro Gln Ile
1055 1060 1065
Cys Gly Pro Gly Thr Trp Leu Thr Pro Phe Leu Gln Ala Val Tyr
1070 1075 1080
Leu Phe Val Gln Tyr Ile Ile Met Val Asn Leu Leu Ile Ala Phe
1085 1090 1095
Phe Asn Asn Val Tyr Leu Gln Val Lys Ala Ile Ser Asn Ile Val
1100 1105 1110
Trp Lys Tyr Gln Arg Tyr His Phe Ile Met Ala Tyr His Glu Lys
1115 1120 1125
Pro Val Leu Pro Pro Pro Leu Ile Ile Leu Ser His Ile Val Ser
1130 1135 1140
Leu Phe Cys Cys Ile Cys Lys Arg Arg Lys Lys Asp Lys Thr Ser
1145 1150 1155
Asp Gly Pro Lys Leu Phe Leu Thr Glu Glu Asp Gln Lys Lys Leu
1160 1165 1170
His Asp Phe Glu Glu Gln Cys Val Glu Met Tyr Phe Asn Glu Lys
1175 1180 1185
Asp Asp Lys Phe His Ser Gly Ser Glu Glu Arg Ile Arg Val Thr
1190 1195 1200
Phe Glu Arg Val Glu Gln Met Cys Ile Gln Ile Lys Glu Val Gly
1205 1210 1215
Asp Arg Val Asn Tyr Ile Lys Arg Ser Leu Gln Ser Leu Asp Ser
1220 1225 1230
Gln Ile Gly His Leu Gln Asp Leu Ser Ala Leu Thr Val Asp Thr
1235 1240 1245
Leu Lys Thr Leu Thr Ala Gln Lys Ala Ser Glu Ala Ser Lys Val
1250 1255 1260
His Asn Glu Ile Thr Arg Glu Leu Ser Ile Ser Lys His Leu Ala
1265 1270 1275
Gln Asn Leu Ile Asp Asp Gly Pro Val Arg Pro Ser Val Trp Lys
1280 1285 1290
Lys His Gly Val Val Asn Thr Leu Ser Ser Ser Leu Pro Gln Gly
1295 1300 1305
Asp Leu Glu Ser Asn Asn Pro Phe His Cys Asn Ile Leu Met Lys
1310 1315 1320
Asp Asp Lys Asp Pro Gln Cys Asn Ile Phe Gly Gln Asp Leu Pro
1325 1330 1335
Ala Val Pro Gln Arg Lys Glu Phe Asn Phe Pro Glu Ala Gly Ser
1340 1345 1350
Ser Ser Gly Ala Leu Phe Pro Ser Ala Val Ser Pro Pro Glu Leu
1355 1360 1365
Arg Gln Arg Leu His Gly Val Glu Leu Leu Lys Ile Phe Asn Lys
1370 1375 1380
Asn Gln Lys Leu Gly Ser Ser Ser Thr Ser Ile Pro His Leu Ser
1385 1390 1395
Ser Pro Pro Thr Lys Phe Phe Val Ser Thr Pro Ser Gln Pro Ser
1400 1405 1410
Cys Lys Ser His Leu Glu Thr Gly Thr Lys Asp Gln Glu Thr Val
1415 1420 1425
Cys Ser Lys Ala Thr Glu Gly Asp Asn Thr Glu Phe Gly Ala Phe
1430 1435 1440
Val Gly His Arg Asp Ser Met Asp Leu Gln Arg Phe Lys Glu Thr
1445 1450 1455
Ser Asn Lys Ile Lys Ile Leu Ser Asn Asn Asn Thr Ser Glu Asn
1460 1465 1470
Thr Leu Lys Arg Val Ser Ser Leu Ala Gly Phe Thr Asp Cys His
1475 1480 1485
Arg Thr Ser Ile Pro Val His Ser Lys Gln Glu Lys Ile Ser Arg

```
                1490                1495                1500
Arg Pro Ser Thr Glu Asp Thr His Glu Val Asp Ser Lys Ala Ala
                1505                1510                1515
Leu Ile Pro Asp Trp Leu Gln Asp Arg Pro Ser Asn Arg Glu Met
                1520                1525                1530
Pro Ser Glu Glu Gly Thr Leu Asn Gly Leu Thr Ser Pro Phe Lys
                1535                1540                1545
Pro Ala Met Asp Thr Asn Tyr Tyr Tyr Ser Ala Val Glu Arg Asn
                1550                1555                1560
Asn Leu Met Arg Leu Ser Gln Ser Ile Pro Phe Thr Pro Val Pro
                1565                1570                1575
Pro Arg Gly Glu Pro Val Thr Val Tyr Arg Leu Glu Glu Ser Ser
                1580                1585                1590
Pro Asn Ile Leu Asn Asn Ser Met Ser Ser Trp Ser Gln Leu Gly
                1595                1600                1605
Leu Cys Ala Lys Ile Glu Phe Leu Ser Lys Glu Glu Met Gly Gly
                1610                1615                1620
Gly Leu Arg Arg Ala Val Lys Val Gln Cys Thr Trp Ser Glu His
                1625                1630                1635
Asp Ile Leu Lys Ser Gly His Leu Tyr Ile Ile Lys Ser Phe Leu
                1640                1645                1650
Pro Glu Val Val Asn Thr Trp Ser Ser Ile Tyr Lys Glu Asp Thr
                1655                1660                1665
Val Leu His Leu Cys Leu Arg Glu Ile Gln Gln Gln Arg Ala Ala
                1670                1675                1680
Gln Lys Leu Thr Phe Ala Phe Asn Gln Met Lys Pro Lys Ser Ile
                1685                1690                1695
Pro Tyr Ser Pro Arg Phe Leu Glu Val Phe Leu Leu Tyr Cys His
                1700                1705                1710
Ser Ala Gly Gln Trp Phe Ala Val Glu Glu Cys Met Thr Gly Glu
                1715                1720                1725
Phe Arg Lys Tyr Asn Asn Asn Asn Gly Asp Glu Ile Ile Pro Thr
                1730                1735                1740
Asn Thr Leu Glu Glu Ile Met Leu Ala Phe Ser His Trp Thr Tyr
                1745                1750                1755
Glu Tyr Thr Arg Gly Glu Leu Leu Val Leu Asp Leu Gln Gly Val
                1760                1765                1770
Gly Glu Asn Leu Thr Asp Pro Ser Val Ile Lys Ala Glu Glu Lys
                1775                1780                1785
Arg Ser Cys Asp Met Val Phe Gly Pro Ala Asn Leu Gly Glu Asp
                1790                1795                1800
Ala Ile Lys Asn Phe Arg Ala Lys His His Cys Asn Ser Cys Cys
                1805                1810                1815
Arg Lys Leu Lys Leu Pro Asp Leu Lys Arg Asn Asp Tyr Thr Pro
                1820                1825                1830
Asp Lys Ile Ile Phe Pro Gln Asp Glu Pro Ser Asp Leu Asn Leu
                1835                1840                1845
Gln Pro Gly Asn Ser Thr Lys Glu Ser Glu Ser Thr Asn Ser Val
                1850                1855                1860
Arg Leu Met Leu
```

<210> 24
<211> 1237
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 6176128CD1

<400> 24
Met Ala Arg Ala Lys Leu Pro Arg Ser Pro Ser Glu Gly Lys Ala

1 5 10 15
Gly Pro Gly Gly Ala Pro Ala Gly Ala Ala Ala Pro Glu Glu Pro
20 25 30
His Gly Leu Ser Pro Leu Leu Pro Ala Arg Gly Gly Gly Ser Val
35 40 45
Gly Ser Asp Val Gly Gln Arg Leu Pro Val Glu Asp Phe Ser Leu
50 55 60
Asp Ser Ser Leu Ser Gln Val Gln Val Glu Phe Tyr Val Asn Glu
65 70 75
Asn Thr Phe Lys Glu Arg Leu Lys Leu Phe Phe Ile Lys Asn Gln
80 85 90
Arg Ser Ser Leu Arg Ile Arg Leu Phe Asn Phe Ser Leu Lys Leu
95 100 105
Leu Thr Cys Leu Leu Tyr Ile Val Arg Val Leu Leu Asp Asp Pro
110 115 120
Ala Leu Gly Ile Gly Trp Trp Gly Cys Pro Arg Gln Asn Tyr Ser
125 130 135
Phe Asn Asp Ser Ser Ser Glu Ile Asn Trp Ala Pro Ile Leu Trp
140 145 150
Val Glu Arg Lys Met Thr Leu Trp Ala Ile Gln Val Ile Val Ala
155 160 165
Ile Ile Ser Phe Leu Glu Thr Met Leu Leu Ile Tyr Leu Ser Tyr
170 175 180
Lys Gly Asn Ile Trp Glu Gln Ile Phe Arg Val Ser Phe Val Leu
185 190 195
Glu Met Ile Asn Thr Leu Pro Phe Ile Ile Thr Ile Phe Trp Pro
200 205 210
Pro Leu Arg Asn Leu Phe Ile Pro Val Phe Leu Asn Cys Trp Leu
215 220 225
Ala Lys His Ala Leu Glu Asn Met Ile Asn Asp Phe His Arg Ala
230 235 240
Ile Leu Arg Thr Gln Ser Ala Met Phe Asn Gln Val Leu Ile Leu
245 250 255
Phe Cys Thr Leu Leu Cys Leu Val Phe Thr Gly Thr Cys Gly Ile
260 265 270
Gln His Leu Glu Arg Ala Gly Glu Asn Leu Ser Leu Leu Thr Ser
275 280 285
Phe Tyr Phe Cys Ile Val Thr Phe Ser Thr Val Gly Tyr Gly Asp
290 295 300
Val Thr Pro Lys Ile Trp Pro Ser Gln Leu Leu Val Val Ile Met
305 310 315
Ile Cys Val Ala Leu Val Val Leu Pro Leu Gln Phe Glu Glu Leu
320 325 330
Val Tyr Leu Trp Met Glu Arg Gln Lys Ser Gly Gly Asn Tyr Ser
335 340 345
Arg His Arg Ala Gln Thr Glu Lys His Val Val Leu Cys Val Ser
350 355 360
Ser Leu Lys Ile Asp Leu Leu Met Asp Phe Leu Asn Glu Phe Tyr
365 370 375
Ala His Pro Arg Leu Gln Asp Tyr Tyr Val Val Ile Leu Cys Pro
380 385 390
Thr Glu Met Asp Val Gln Val Arg Arg Val Leu Gln Ile Pro Leu
395 400 405
Trp Ser Gln Arg Val Ile Tyr Leu Gln Gly Ser Ala Leu Lys Asp
410 415 420
Gln Asp Leu Met Arg Ala Lys Met Asp Asn Gly Glu Ala Cys Phe
425 430 435
Ile Leu Ser Ser Arg Asn Glu Val Asp Arg Thr Ala Ala Asp His
440 445 450
Gln Thr Ile Leu Arg Ala Trp Ala Val Lys Asp Phe Ala Pro Asn
455 460 465
Cys Pro Leu Tyr Val Gln Ile Leu Lys Pro Glu Asn Lys Phe His
470 475 480

Val Lys Phe Ala Asp His Val Val Cys Glu Glu Glu Cys Lys Tyr
485 490 495
Ala Met Leu Ala Leu Asn Cys Ile Cys Pro Ala Thr Ser Thr Leu
500 505 510
Ile Thr Leu Leu Val His Thr Ser Arg Gly Gln Glu Gly Gln Glu
515 520 525
Ser Pro Glu Gln Trp Gln Arg Met Tyr Gly Arg Cys Ser Gly Asn
530 535 540
Glu Val Tyr His Ile Arg Met Gly Asp Ser Lys Phe Phe Arg Glu
545 550 555
Tyr Glu Gly Lys Ser Phe Thr Tyr Ala Ala Phe His Ala His Lys
560 565 570
Lys Tyr Gly Val Cys Leu Ile Gly Leu Lys Arg Glu Asp Asn Lys
575 580 585
Ser Ile Leu Leu Asn Pro Gly Pro Arg His Ile Leu Ala Ala Ser
590 595 600
Asp Thr Cys Phe Tyr Ile Asn Ile Thr Lys Glu Glu Asn Ser Ala
605 610 615
Phe Ile Phe Lys Gln Glu Glu Lys Arg Lys Lys Arg Ala Phe Ser
620 625 630
Gly Gln Gly Leu His Glu Gly Pro Ala Arg Leu Pro Val His Ser
635 640 645
Ile Ile Ala Ser Met Gly Thr Val Ala Met Asp Leu Gln Gly Thr
650 655 660
Glu His Arg Pro Thr Gln Ser Gly Gly Gly Gly Gly Gly Ser Lys
665 670 675
Leu Ala Leu Pro Thr Glu Asn Gly Ser Gly Ser Arg Arg Pro Ser
680 685 690
Ile Ala Pro Val Leu Glu Leu Ala Asp Ser Ser Ala Leu Leu Pro
695 700 705
Cys Asp Leu Leu Ser Asp Gln Ser Glu Asp Glu Val Thr Pro Ser
710 715 720
Asp Asp Glu Gly Leu Ser Val Val Glu Tyr Val Lys Gly Tyr Pro
725 730 735
Pro Asn Ser Pro Tyr Ile Gly Ser Ser Pro Thr Leu Cys His Leu
740 745 750
Leu Pro Val Lys Ala Pro Phe Cys Cys Leu Arg Leu Asp Lys Gly
755 760 765
Cys Lys His Asn Ser Tyr Glu Asp Ala Lys Ala Tyr Gly Phe Lys
770 775 780
Asn Lys Leu Ile Ile Val Ser Ala Glu Thr Ala Gly Asn Gly Leu
785 790 795
Tyr Asn Phe Ile Val Pro Leu Arg Ala Tyr Tyr Arg Ser Arg Lys
800 805 810
Glu Leu Asn Pro Ile Val Leu Leu Leu Asp Asn Lys Pro Asp His
815 820 825
His Phe Leu Glu Ala Ile Cys Cys Phe Pro Met Val Tyr Tyr Met
830 835 840
Glu Gly Ser Val Asp Asn Leu Asp Ser Leu Leu Gln Cys Gly Ile
845 850 855
Ile Tyr Ala Asp Asn Leu Val Val Val Asp Lys Glu Ser Thr Met
860 865 870
Ser Ala Glu Glu Asp Tyr Met Ala Asp Ala Lys Thr Ile Val Asn
875 880 885
Val Gln Thr Met Phe Arg Leu Phe Pro Ser Leu Ser Ile Thr Thr
890 895 900
Glu Leu Thr His Pro Ser Asn Met Arg Phe Met Gln Phe Arg Ala
905 910 915
Lys Asp Ser Tyr Ser Leu Ala Leu Ser Lys Leu Glu Lys Arg Glu
920 925 930
Arg Glu Asn Gly Ser Asn Leu Ala Phe Met Phe Arg Leu Pro Phe
935 940 945
Ala Ala Gly Arg Val Phe Ser Ile Ser Met Leu Asp Thr Leu Leu

950 955 960
Tyr Gln Ser Phe Val Lys Asp Tyr Met Ile Thr Ile Thr Arg Leu
965 970 975
Leu Leu Gly Leu Asp Thr Thr Pro Gly Ser Gly Tyr Leu Cys Ala
980 985 990
Met Lys Ile Thr Glu Gly Asp Leu Trp Ile Arg Thr Tyr Gly Arg
995 1000 1005
Leu Phe Gln Lys Leu Cys Ser Ser Ser Ala Glu Ile Pro Ile Gly
1010 1015 1020
Ile Tyr Arg Thr Glu Ser His Val Phe Ser Thr Ser Glu Pro His
1025 1030 1035
Glu Leu Arg Ala Gln Ser Gln Ile Ser Val Asn Val Glu Asp Cys
1040 1045 1050
Glu Asp Thr Arg Glu Val Lys Gly Pro Trp Gly Ser Arg Ala Gly
1055 1060 1065
Thr Gly Gly Ser Ser Gln Gly Arg His Thr Gly Gly Gly Asp Pro
1070 1075 1080
Ala Glu His Pro Leu Leu Arg Arg Lys Ser Leu Gln Trp Ala Arg
1085 1090 1095
Arg Leu Ser Arg Lys Ala Pro Lys Gln Ala Gly Arg Ala Ala Ala
1100 1105 1110
Ala Glu Trp Ile Ser Gln Gln Arg Leu Ser Leu Tyr Arg Arg Ser
1115 1120 1125
Glu Arg Gln Glu Leu Ser Glu Leu Val Lys Asn Arg Met Lys His
1130 1135 1140
Leu Gly Leu Pro Thr Thr Gly Tyr Glu Asp Val Ala Asn Leu Thr
1145 1150 1155
Ala Ser Asp Val Met Asn Arg Val Asn Leu Gly Tyr Leu Gln Asp
1160 1165 1170
Glu Met Asn Asp His Gln Asn Thr Leu Ser Tyr Val Leu Ile Asn
1175 1180 1185
Pro Pro Pro Asp Thr Arg Leu Glu Pro Ser Asp Ile Val Tyr Leu
1190 1195 1200
Ile Arg Ser Asp Pro Leu Ala His Val Ala Ser Ser Ser Gln Ser
1205 1210 1215
Arg Lys Ser Ser Cys Ser His Lys Leu Ser Ser Cys Asn Pro Glu
1220 1225 1230
Thr Arg Asp Glu Thr Gln Leu
1235

<210> 25
<211> 539
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7473418CD1

<400> 25
Met Ala Ser Ala Leu Ser Tyr Val Ser Lys Phe Lys Ser Phe Val
1 5 10 15
Ile Leu Phe Val Thr Pro Leu Leu Leu Leu Pro Leu Val Ile Leu
20 25 30
Met Pro Ala Lys Phe Val Arg Cys Ala Tyr Val Ile Ile Leu Met
35 40 45
Ala Ile Tyr Trp Cys Thr Glu Val Ile Pro Leu Ala Val Thr Ser
50 55 60
Leu Met Pro Val Leu Leu Phe Pro Leu Phe Gln Ile Leu Asp Ser
65 70 75
Arg Gln Val Cys Val Gln Tyr Met Lys Asp Thr Asn Met Leu Phe
80 85 90
Leu Gly Gly Leu Ile Val Ala Val Ala Val Glu Arg Trp Asn Leu

```
                    95                  100                 105
His Lys Arg Ile Ala Leu Arg Thr Leu Leu Trp Val Gly Ala Lys
                    110                 115                 120
Pro Ala Arg Leu Met Leu Gly Phe Met Gly Val Thr Ala Leu Leu
                    125                 130                 135
Ser Met Trp Ile Ser Asn Thr Ala Thr Thr Ala Met Met Val Pro
                    140                 145                 150
Ile Val Glu Ala Ile Leu Gln Gln Met Glu Ala Thr Ser Ala Ala
                    155                 160                 165
Thr Glu Ala Gly Leu Glu Leu Val Asp Lys Gly Lys Ala Lys Glu
                    170                 175                 180
Leu Pro Ala Asn Ser Ala Val Pro Thr Thr Gly Ser Gln Val Ile
                    185                 190                 195
Phe Glu Gly Pro Thr Leu Gly Gln Gln Glu Asp Gln Glu Arg Lys
                    200                 205                 210
Arg Leu Cys Lys Ala Met Thr Leu Cys Ile Cys Tyr Ala Ala Ser
                    215                 220                 225
Ile Gly Gly Thr Ala Thr Leu Thr Gly Thr Gly Pro Asn Val Val
                    230                 235                 240
Leu Leu Gly Gln Met Asn Glu Leu Phe Pro Asp Ser Lys Asp Leu
                    245                 250                 255
Val Asn Phe Ala Ser Trp Phe Ala Phe Ala Phe Pro Asn Met Leu
                    260                 265                 270
Val Met Leu Leu Phe Ala Trp Leu Trp Leu Gln Phe Val Tyr Met
                    275                 280                 285
Arg Phe Asn Phe Lys Lys Ser Trp Gly Cys Gly Leu Glu Ser Lys
                    290                 295                 300
Lys Asn Glu Lys Ala Ala Leu Lys Val Leu Gln Glu Glu Tyr Arg
                    305                 310                 315
Lys Leu Gly Pro Leu Ser Phe Ala Glu Ile Asn Val Leu Ile Cys
                    320                 325                 330
Phe Phe Leu Leu Val Ile Leu Trp Phe Ser Arg Asp Pro Gly Phe
                    335                 340                 345
Met Pro Gly Trp Leu Thr Val Ala Trp Val Glu Glu Arg Lys Thr
                    350                 355                 360
Pro Phe Tyr Pro Pro Pro Leu Leu Asp Trp Lys Val Thr Gln Glu
                    365                 370                 375
Lys Val Pro Trp Gly Ile Val Leu Leu Leu Gly Gly Gly Phe Ala
                    380                 385                 390
Leu Ala Lys Gly Ser Glu Ala Ser Gly Leu Ser Val Trp Met Gly
                    395                 400                 405
Lys Gln Met Glu Pro Leu His Ala Val Pro Pro Ala Ala Ile Thr
                    410                 415                 420
Leu Ile Leu Ser Leu Leu Val Ala Val Phe Thr Glu Cys Thr Ser
                    425                 430                 435
Asn Val Ala Thr Thr Thr Leu Phe Leu Pro Ile Phe Ala Ser Met
                    440                 445                 450
Ser Arg Ser Ile Gly Leu Asn Pro Leu Tyr Ile Met Leu Pro Cys
                    455                 460                 465
Thr Leu Ser Ala Ser Phe Ala Phe Met Leu Pro Val Ala Thr Pro
                    470                 475                 480
Pro Asn Ala Ile Val Phe Thr Tyr Gly His Leu Lys Val Ala Asp
                    485                 490                 495
Met Val Lys Thr Gly Val Ile Met Asn Ile Ile Gly Val Phe Cys
                    500                 505                 510
Val Phe Leu Ala Val Asn Thr Trp Gly Arg Ala Ile Phe Asp Leu
                    515                 520                 525
Asp His Phe Pro Asp Trp Ala Asn Val Thr His Ile Glu Thr
                    530                 535
```

<210> 26
<211> 755
<212> PRT

<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7474129CD1

<400> 26

```
Met Lys Ala His Pro Lys Glu Met Val Pro Leu Met Gly Lys Arg
 1               5                  10                  15
Val Ala Ala Pro Ser Gly Asn Pro Ala Val Leu Pro Glu Lys Arg
                20                  25                  30
Pro Ala Glu Ile Thr Pro Thr Lys Lys Ser Ile Ser Gly Asn Cys
                35                  40                  45
Asp Asp Met Asp Ser Pro Gln Ser Pro Gln Asp Asp Val Thr Glu
                50                  55                  60
Thr Pro Ser Asn Pro Asn Ser Pro Ser Ala Gln Leu Ala Lys Glu
                65                  70                  75
Glu Gln Arg Arg Lys Lys Arg Arg Leu Lys Lys Arg Ile Phe Ala
                80                  85                  90
Ala Val Ser Glu Gly Cys Val Glu Glu Leu Val Glu Leu Leu Val
                95                 100                 105
Glu Leu Gln Glu Leu Cys Arg Arg Arg His Asp Glu Asp Val Pro
               110                 115                 120
Asp Phe Leu Met His Lys Leu Thr Ala Ser Asp Thr Gly Lys Thr
               125                 130                 135
Cys Leu Met Lys Ala Leu Leu Asn Ile Asn Pro Asn Thr Lys Glu
               140                 145                 150
Ile Val Arg Ile Leu Leu Ala Phe Ala Glu Glu Asn Asp Ile Leu
               155                 160                 165
Gly Arg Phe Ile Asn Ala Glu Tyr Thr Glu Glu Ala Tyr Glu Gly
               170                 175                 180
Gln Thr Ala Leu Asn Ile Ala Ile Glu Arg Arg Gln Gly Asp Ile
               185                 190                 195
Ala Ala Leu Leu Ile Ala Ala Gly Ala Asp Val Asn Ala His Ala
               200                 205                 210
Lys Gly Ala Phe Phe Asn Pro Lys Tyr Gln His Glu Gly Phe Tyr
               215                 220                 225
Phe Gly Glu Thr Pro Leu Ala Leu Ala Ala Cys Thr Asn Gln Pro
               230                 235                 240
Glu Ile Val Gln Leu Leu Met Glu His Glu Gln Thr Asp Ile Thr
               245                 250                 255
Ser Arg Asp Ser Arg Gly Asn Asn Ile Leu His Ala Leu Val Thr
               260                 265                 270
Val Ala Glu Asp Phe Lys Thr Gln Asn Asp Val Val Lys Arg Met
               275                 280                 285
Tyr Asp Met Ile Leu Leu Arg Ser Gly Asn Trp Glu Leu Glu Thr
               290                 295                 300
Thr Arg Asn Asn Asp Gly Leu Thr Pro Leu Gln Leu Ala Ala Lys
               305                 310                 315
Met Gly Lys Ala Glu Ile Leu Lys Tyr Ile Leu Ser Arg Glu Ile
               320                 325                 330
Lys Glu Lys Arg Leu Arg Ser Leu Ser Arg Lys Phe Thr Asp Trp
               335                 340                 345
Ala Tyr Gly Pro Val Ser Ser Ser Leu Tyr Asp Leu Thr Asn Val
               350                 355                 360
Asp Thr Thr Thr Asp Asn Ser Val Leu Glu Ile Thr Val Tyr Asn
               365                 370                 375
Thr Asn Ile Asp Asn Arg His Glu Met Leu Thr Leu Glu Pro Leu
               380                 385                 390
His Thr Leu Leu His Met Lys Trp Lys Lys Phe Ala Lys His Met
               395                 400                 405
Phe Phe Leu Ser Phe Cys Phe Tyr Phe Phe Tyr Asn Ile Thr Leu
               410                 415                 420
```

```
Thr Leu Val Ser Tyr Tyr Arg Pro Arg Glu Glu Glu Ala Ile Pro
                425                 430                 435
His Pro Leu Ala Leu Thr His Lys Met Gly Trp Leu Gln Leu Leu
                440                 445                 450
Gly Arg Met Phe Val Leu Ile Trp Ala Met Cys Ile Ser Val Lys
                455                 460                 465
Glu Gly Ile Ala Ile Phe Leu Leu Arg Pro Ser Asp Leu Gln Ser
                470                 475                 480
Ile Leu Ser Asp Ala Trp Phe His Phe Val Phe Phe Ile Gln Ala
                485                 490                 495
Val Leu Val Ile Leu Ser Val Phe Leu Tyr Leu Phe Ala Tyr Lys
                500                 505                 510
Glu Tyr Leu Ala Cys Leu Val Leu Ala Met Ala Leu Gly Trp Ala
                515                 520                 525
Asn Met Leu Tyr Tyr Thr Arg Gly Phe Gln Ser Met Gly Met Tyr
                530                 535                 540
Ser Val Met Ile Gln Lys Val Ile Leu His Asp Val Leu Lys Phe
                545                 550                 555
Leu Phe Val Tyr Ile Val Phe Leu Leu Gly Phe Gly Val Ala Leu
                560                 565                 570
Ala Ser Leu Ile Glu Lys Cys Pro Lys Asp Asn Lys Asp Cys Ser
                575                 580                 585
Ser Tyr Gly Ser Phe Ser Asp Ala Val Leu Glu Leu Phe Lys Leu
                590                 595                 600
Thr Ile Gly Leu Gly Asp Leu Asn Ile Gln Gln Asn Ser Lys Tyr
                605                 610                 615
Pro Ile Leu Phe Leu Phe Leu Leu Ile Thr Tyr Val Ile Leu Thr
                620                 625                 630
Phe Val Leu Leu Leu Asn Met Leu Ile Ala Leu Met Gly Glu Thr
                635                 640                 645
Val Glu Asn Val Ser Lys Glu Ser Glu Arg Ile Trp Arg Leu Gln
                650                 655                 660
Arg Ala Arg Thr Ile Leu Glu Phe Glu Lys Met Leu Pro Glu Trp
                665                 670                 675
Leu Arg Ser Arg Phe Arg Met Gly Glu Leu Cys Lys Val Ala Glu
                680                 685                 690
Asp Asp Phe Arg Leu Cys Leu Arg Ile Asn Glu Val Lys Trp Thr
                695                 700                 705
Glu Trp Lys Thr His Val Ser Phe Leu Asn Glu Asp Pro Gly Pro
                710                 715                 720
Val Arg Arg Thr Asp Phe Asn Lys Ile Gln Asp Ser Ser Arg Asn
                725                 730                 735
Asn Ser Lys Thr Thr Leu Asn Ala Phe Glu Glu Val Glu Glu Phe
                740                 745                 750
Pro Glu Thr Ser Val
                755
```

<210> 27
<211> 301
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7481414CD1

<400> 27

```
Met Lys Ser His Pro Ala Ile Gln Ala Ala Ile Asp Leu Thr Ala
  1               5                  10                  15
Gly Ala Ala Gly Gly Gly Ala Cys Val Leu Thr Gly Gln Pro Phe
                 20                  25                  30
Asp Thr Ile Lys Val Lys Met Gln Thr Phe Pro Gln Leu Tyr Lys
                 35                  40                  45
```

```
Gly Leu Ala Asp Cys Phe Leu Lys Thr Tyr Asn Gln Val Gly Ile
                50                  55                  60
Arg Gly Leu Tyr Arg Gly Thr Ser Pro Ala Leu Leu Ala Tyr Val
                65                  70                  75
Thr Gln Gly Ser Val Leu Phe Met Cys Phe Gly Phe Cys Gln Gln
                80                  85                  90
Phe Val Arg Lys Val Ala Arg Val Glu Gln Asn Ala Glu Leu Asn
                95                  100                 105
Asp Leu Glu Thr Ala Thr Ala Gly Ser Leu Ala Ser Ala Phe Ala
                110                 115                 120
Ala Leu Ala Leu Cys Pro Thr Glu Leu Val Lys Cys Arg Leu Gln
                125                 130                 135
Thr Met Tyr Glu Met Lys Met Ser Gly Lys Ile Ala Gln Ser Tyr
                140                 145                 150
Asn Thr Ile Trp Ser Met Val Lys Ser Ile Phe Met Lys Asp Gly
                155                 160                 165
Pro Leu Gly Phe Tyr Arg Gly Leu Ser Thr Thr Leu Ala Gln Glu
                170                 175                 180
Ile Pro Gly Tyr Phe Phe Tyr Phe Gly Gly Tyr Glu Ile Ser Arg
                185                 190                 195
Ser Phe Phe Ala Ser Gly Gly Ser Lys Asp Glu Leu Gly Pro Val
                200                 205                 210
Pro Leu Met Leu Ser Gly Gly Phe Ala Gly Ile Cys Leu Trp Leu
                215                 220                 225
Ile Ile Phe Pro Val Asp Cys Ile Lys Ser Arg Ile Gln Val Leu
                230                 235                 240
Ser Met Phe Gly Lys Pro Ala Gly Leu Ile Glu Thr Phe Ile Ser
                245                 250                 255
Val Val Arg Asn Glu Gly Ile Ser Ala Leu Tyr Ser Gly Leu Lys
                260                 265                 270
Ala Thr Leu Ile Arg Ala Ile Pro Ser Asn Ala Ala Leu Phe Leu
                275                 280                 285
Val Tyr Glu Tyr Ser Arg Lys Met Met Met Asn Met Val Glu Glu
                290                 295                 300
Tyr


<210> 28
<211> 515
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7481461CD1

<400> 28
Met Val Leu Ser Gln Glu Glu Pro Asp Ser Ala Arg Gly Thr Ser
 1              5                   10                  15
Glu Ala Gln Pro Leu Gly Pro Ala Pro Thr Gly Ala Ala Pro Pro
                20                  25                  30
Pro Gly Pro Gly Pro Ser Asp Ser Pro Glu Ala Ala Val Glu Lys
                35                  40                  45
Val Glu Val Glu Leu Ala Gly Pro Ala Thr Ala Glu Pro His Glu
                50                  55                  60
Pro Pro Glu Pro Pro Glu Gly Gly Trp Gly Trp Leu Val Met Leu
                65                  70                  75
Ala Ala Met Trp Cys Asn Gly Ser Val Phe Gly Ile Gln Asn Ala
                80                  85                  90
Cys Gly Val Leu Phe Val Ser Met Leu Glu Thr Phe Gly Ser Lys
                95                  100                 105
Asp Asp Asp Lys Met Val Phe Lys Thr Ala Trp Val Gly Ser Leu
                110                 115                 120
```

Ser Met Gly Met Ile Phe Phe Cys Cys Pro Ile Val Ser Val Phe
125 130 135
Thr Asp Leu Phe Gly Cys Arg Lys Thr Ala Val Val Gly Ala Ala
140 145 150
Val Gly Phe Val Gly Leu Met Ser Ser Ser Phe Val Ser Ser Ile
155 160 165
Glu Pro Leu Tyr Leu Thr Tyr Gly Ile Ile Phe Ala Cys Gly Cys
170 175 180
Ser Phe Ala Tyr Gln Pro Ser Leu Val Ile Leu Gly His Tyr Phe
185 190 195
Lys Lys Arg Leu Gly Leu Val Asn Gly Ile Val Thr Ala Gly Ser
200 205 210
Ser Val Phe Thr Ile Leu Leu Pro Leu Leu Leu Arg Val Leu Ile
215 220 225
Asp Ser Val Gly Leu Phe Tyr Thr Leu Arg Val Leu Cys Ile Phe
230 235 240
Met Phe Val Leu Phe Leu Ala Gly Phe Thr Tyr Arg Pro Leu Ala
245 250 255
Thr Ser Thr Lys Asp Lys Glu Ser Gly Gly Ser Gly Ser Ser Leu
260 265 270
Phe Ser Arg Lys Lys Phe Ser Pro Pro Lys Lys Ile Phe Asn Phe
275 280 285
Ala Ile Phe Lys Val Thr Ala Tyr Ala Val Trp Ala Val Gly Ile
290 295 300
Pro Leu Ala Leu Phe Gly Tyr Phe Val Pro Tyr Val His Leu Met
305 310 315
Lys His Val Asn Glu Arg Phe Gln Asp Glu Lys Asn Lys Glu Val
320 325 330
Val Leu Met Cys Ile Gly Val Thr Ser Gly Val Gly Arg Leu Leu
335 340 345
Phe Gly Arg Ile Ala Asp Tyr Val Pro Gly Val Lys Lys Val Tyr
350 355 360
Leu Gln Val Leu Ser Phe Phe Phe Ile Gly Leu Met Ser Met Met
365 370 375
Ile Pro Leu Cys Ser Ile Phe Gly Ala Leu Ile Ala Val Cys Leu
380 385 390
Ile Met Gly Leu Phe Asp Gly Cys Phe Ile Ser Ile Met Ala Pro
395 400 405
Ile Ala Phe Glu Leu Val Gly Ala Gln Asp Val Ser Gln Ala Ile
410 415 420
Gly Phe Leu Leu Gly Phe Met Ser Ile Pro Met Thr Val Gly Pro
425 430 435
Pro Ile Ala Gly Leu Leu Arg Asp Lys Leu Gly Ser Tyr Asp Val
440 445 450
Ala Phe Tyr Leu Ala Gly Val Pro Pro Leu Ile Gly Gly Ala Val
455 460 465
Leu Cys Phe Ile Pro Trp Ile His Ser Lys Lys Gln Arg Glu Ile
470 475 480
Ser Lys Thr Thr Gly Lys Glu Lys Met Glu Lys Met Leu Glu Asn
485 490 495
Gln Asn Ser Leu Leu Ser Ser Ser Ser Gly Met Phe Lys Lys Glu
500 505 510
Ser Asp Ser Ile Ile
515

<210> 29
<211> 1519
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7472541CD1

<400> 29
Met Ala Leu Ser Val Asp Ser Ser Trp His Arg Trp Gln Trp Arg
1 5 10 15
Val Arg Asp Gly Phe Pro His Cys Pro Ser Glu Thr Thr Pro Leu
20 25 30
Leu Ser Pro Glu Lys Gly Arg Gln Ser Tyr Asn Leu Thr Gln Gln
35 40 45
Arg Val Val Phe Pro Asn Asn Ser Ile Phe His Gln Asp Trp Glu
50 55 60
Glu Val Ser Arg Arg Tyr Pro Gly Asn Arg Thr Cys Thr Thr Lys
65 70 75
Tyr Thr Leu Phe Thr Phe Leu Pro Arg Asn Leu Phe Glu Gln Phe
80 85 90
His Arg Trp Ala Asn Leu Tyr Phe Leu Phe Leu Val Ile Leu Ser
95 100 105
Trp Met Pro Ser Met Glu Val Phe His Arg Glu Ile Thr Met Leu
110 115 120
Pro Leu Ala Ile Val Leu Phe Val Ile Met Ile Lys Asp Gly Met
125 130 135
Glu Asp Phe Lys Arg His Arg Phe Asp Lys Ala Ile Asn Cys Ser
140 145 150
Asn Ile Arg Ile Tyr Glu Arg Lys Glu Gln Thr Tyr Val Gln Lys
155 160 165
Cys Trp Lys Asp Val Arg Val Gly Asp Phe Ile Gln Met Lys Cys
170 175 180
Asn Glu Ile Val Pro Ala Asp Ile Leu Leu Leu Phe Ser Ser Asp
185 190 195
Pro Asn Gly Ile Cys His Leu Glu Thr Ala Ser Leu Asp Gly Glu
200 205 210
Thr Asn Leu Lys Gln Arg Arg Val Val Lys Gly Phe Ser Gln Gln
215 220 225
Glu Val Gln Phe Glu Pro Glu Leu Phe His Asn Thr Ile Val Cys
230 235 240
Glu Lys Pro Asn Asn His Leu Asn Lys Phe Lys Gly Tyr Met Glu
245 250 255
His Pro Asp Gln Thr Arg Thr Gly Phe Gly Cys Glu Ser Leu Leu
260 265 270
Leu Arg Gly Cys Thr Ile Arg Asn Thr Glu Met Ala Val Gly Ile
275 280 285
Val Ile Tyr Ala Gly His Glu Thr Lys Ala Met Leu Asn Asn Ser
290 295 300
Gly Pro Arg Tyr Lys Arg Ser Lys Ile Glu Arg Arg Met Asn Ile
305 310 315
Asp Ile Phe Phe Cys Ile Gly Ile Leu Ile Leu Met Cys Leu Ile
320 325 330
Gly Ala Val Gly His Ser Ile Trp Asn Gly Thr Phe Glu Glu His
335 340 345
Pro Pro Phe Asp Val Pro Asp Ala Asn Gly Ser Phe Leu Pro Ser
350 355 360
Ala Leu Gly Gly Phe Tyr Met Phe Leu Thr Met Ile Ile Leu Leu
365 370 375
Gln Val Leu Ile Pro Ile Ser Leu Tyr Val Ser Ile Glu Leu Val
380 385 390
Lys Leu Gly Gln Val Phe Phe Leu Ser Asn Asp Leu Asp Leu Tyr
395 400 405
Asp Glu Glu Thr Asp Leu Ser Ile Gln Cys Arg Ala Leu Asn Ile
410 415 420
Ala Glu Asp Leu Gly Gln Ile Gln Tyr Ile Phe Ser Asp Lys Thr
425 430 435
Gly Thr Leu Thr Glu Asn Lys Met Val Phe Arg Arg Cys Thr Ile
440 445 450
Met Gly Ser Glu Tyr Ser His Gln Glu Asn Ala Lys Arg Leu Glu
455 460 465

Thr Pro Lys Glu Leu Asp Ser Asp Gly Glu Glu Trp Thr Gln Tyr
470 475 480
Gln Cys Leu Ser Phe Ser Ala Arg Trp Ala Gln Asp Pro Ala Thr
485 490 495
Met Arg Ser Gln Lys Gly Ala Gln Pro Leu Arg Arg Ser Gln Ser
500 505 510
Ala Arg Val Pro Ile Gln Gly His Tyr Arg Gln Arg Ser Met Gly
515 520 525
His Arg Glu Ser Ser Gln Pro Pro Val Ala Phe Ser Ser Ser Ile
530 535 540
Glu Lys Asp Val Thr Pro Asp Lys Asn Leu Leu Thr Lys Val Arg
545 550 555
Asp Ala Ala Leu Trp Leu Glu Thr Leu Ser Asp Ser Arg Pro Ala
560 565 570
Lys Ala Ser Leu Ser Thr Thr Ser Ser Ile Ala Asp Phe Phe Leu
575 580 585
Ala Leu Thr Ile Cys Asn Ser Val Met Val Ser Thr Thr Thr Glu
590 595 600
Pro Arg Gln Arg Trp Asp Asp Gln Lys Ile Val Glu Asn Asp His
605 610 615
Cys Gln Cys Leu Glu Phe Gln Gly Trp Arg Lys Ile Ser Gly Phe
620 625 630
Thr Tyr Cys Lys Ser Thr Phe Ile Phe Arg Ile Arg Gln Leu Gly
635 640 645
Ile Ile Ser Asn Ile Glu Ser Asn Ile Pro Leu Ser Phe Phe Gly
650 655 660
His Lys Val Thr Ile Lys Pro Ser Ser Lys Ala Leu Gly Thr Ser
665 670 675
Leu Glu Lys Ile Gln Gln Leu Phe Gln Lys Leu Lys Leu Leu Ser
680 685 690
Leu Ser Gln Ser Phe Ser Ser Thr Ala Pro Ser Asp Thr Asp Leu
695 700 705
Gly Glu Ser Leu Gly Ala Asn Val Ala Thr Thr Asp Ser Asp Glu
710 715 720
Arg Asp Asp Ala Ser Val Cys Ser Gly Gly Asp Ser Thr Asp Asp
725 730 735
Gly Gly Tyr Arg Ser Ser Met Trp Asp Gln Gly Asp Ile Leu Glu
740 745 750
Ser Gly Ser Gly Thr Ser Leu Glu Glu Ala Leu Glu Ala Pro Ala
755 760 765
Thr Asp Leu Ala Arg Pro Glu Phe Cys Tyr Glu Ala Glu Ser Pro
770 775 780
Asp Glu Ala Ala Leu Val His Ala Ala His Ala Tyr Ser Phe Thr
785 790 795
Leu Val Ser Arg Thr Pro Glu Gln Val Thr Val Arg Leu Pro Gln
800 805 810
Gly Thr Cys Leu Thr Phe Ser Leu Leu Cys Thr Leu Gly Phe Asp
815 820 825
Ser Val Arg Lys Arg Met Ser Val Val Val Arg His Pro Leu Thr
830 835 840
Gly Glu Ile Val Val Tyr Thr Lys Gly Ala Asp Ser Val Ile Met
845 850 855
Asp Leu Leu Glu Asp Pro Ala Cys Val Pro Asp Ile Asn Met Glu
860 865 870
Lys Lys Leu Arg Lys Ile Arg Ala Arg Thr Gln Lys His Leu Asp
875 880 885
Leu Tyr Ala Arg Asp Gly Leu Arg Thr Leu Cys Ile Ala Lys Lys
890 895 900
Val Val Ser Glu Glu Asp Phe Arg Arg Trp Ala Ser Phe Arg Arg
905 910 915
Glu Ala Glu Ala Ser Leu Asp Asn Arg Asp Glu Leu Leu Met Glu
920 925 930
Thr Ala Gln His Leu Glu Asn Gln Leu Thr Leu Leu Gly Ala Thr

935 940 945
Gly Ile Glu Asp Arg Leu Gln Glu Gly Val Pro Asp Thr Ile Ala
950 955 960
Thr Leu Arg Glu Ala Gly Ile Gln Leu Trp Val Leu Thr Gly Asp
965 970 975
Lys Gln Glu Thr Ala Val Asn Ile Ala His Ser Cys Arg Leu Leu
980 985 990
Asn Gln Thr Asp Thr Val Tyr Thr Ile Asn Thr Glu Asn Gln Glu
995 1000 1005
Thr Cys Glu Ser Ile Leu Asn Cys Ala Leu Glu Glu Leu Lys Gln
1010 1015 1020
Phe Arg Glu Leu Gln Lys Pro Asp Arg Lys Leu Phe Gly Phe Arg
1025 1030 1035
Leu Pro Ser Lys Thr Pro Ser Ile Thr Ser Glu Ala Val Val Pro
1040 1045 1050
Glu Ala Gly Leu Val Ile Asp Gly Lys Thr Leu Asn Ala Ile Phe
1055 1060 1065
Gln Gly Lys Leu Glu Lys Lys Phe Leu Glu Leu Thr Gln Tyr Cys
1070 1075 1080
Arg Ser Val Leu Cys Cys Arg Ser Thr Pro Leu Gln Lys Ser Met
1085 1090 1095
Ile Val Lys Leu Val Arg Asp Lys Leu Arg Val Met Thr Leu Ser
1100 1105 1110
Ile Gly Asp Gly Ala Asn Asp Val Ser Met Ile Gln Ala Ala Asp
1115 1120 1125
Ile Gly Ile Gly Ile Ser Gly Gln Glu Gly Met Gln Ala Val Met
1130 1135 1140
Ser Ser Asp Phe Ala Ile Thr Arg Phe Lys His Leu Lys Lys Leu
1145 1150 1155
Leu Leu Val His Gly His Trp Cys Tyr Ser Arg Leu Ala Arg Met
1160 1165 1170
Val Val Tyr Tyr Leu Tyr Lys Asn Val Cys Tyr Val Asn Leu Leu
1175 1180 1185
Phe Trp Tyr Gln Phe Phe Cys Gly Phe Ser Ser Ser Thr Met Ile
1190 1195 1200
Asp Tyr Trp Gln Met Ile Phe Phe Asn Leu Phe Phe Thr Ser Leu
1205 1210 1215
Pro Pro Leu Val Phe Gly Val Leu Asp Lys Asp Ile Ser Ala Glu
1220 1225 1230
Thr Leu Leu Ala Leu Pro Glu Leu Tyr Lys Ser Gly Gln Asn Ser
1235 1240 1245
Glu Cys Tyr Asn Leu Ser Thr Phe Trp Ile Ser Met Val Asp Ala
1250 1255 1260
Phe Tyr Gln Ser Leu Ile Cys Phe Phe Ile Pro Tyr Leu Ala Tyr
1265 1270 1275
Lys Gly Ser Asp Ile Asp Val Phe Thr Phe Gly Thr Pro Ile Asn
1280 1285 1290
Thr Ile Ser Leu Thr Thr Ile Leu Leu His Gln Ala Met Glu Met
1295 1300 1305
Lys Thr Trp Thr Ile Phe His Gly Val Val Leu Leu Gly Ser Phe
1310 1315 1320
Leu Met Tyr Phe Leu Val Ser Leu Leu Tyr Asn Ala Thr Cys Val
1325 1330 1335
Ile Cys Asn Ser Pro Thr Asn Pro Tyr Trp Val Met Glu Gly Gln
1340 1345 1350
Leu Ser Asn Pro Thr Phe Tyr Leu Val Cys Phe Leu Thr Pro Val
1355 1360 1365
Val Ala Leu Leu Pro Arg Tyr Phe Phe Leu Ser Leu Gln Gly Thr
1370 1375 1380
Cys Gly Lys Ser Leu Ile Ser Lys Ala Gln Lys Ile Asp Lys Leu
1385 1390 1395
Pro Pro Asp Lys Arg Asn Leu Glu Ile Gln Ser Trp Arg Ser Arg
1400 1405 1410

```
Gln Arg Pro Ala Pro Val Pro Glu Val Ala Arg Pro Thr His His
                1415                1420                1425
Pro Val Ser Ser Ile Thr Gly Gln Asp Phe Ser Ala Ser Thr Pro
                1430                1435                1440
Lys Ser Ser Asn Pro Pro Lys Arg Lys His Val Glu Glu Ser Val
                1445                1450                1455
Leu His Glu Gln Arg Cys Gly Thr Glu Cys Met Arg Asp Asp Ser
                1460                1465                1470
Cys Ser Gly Asp Ser Ser Ala Gln Leu Ser Ser Gly Glu His Leu
                1475                1480                1485
Leu Gly Pro Asn Arg Ile Met Ala Tyr Ser Gly Gly Gln Thr Asp
                1490                1495                1500
Met Cys Arg Cys Ser Lys Arg Ser Ser His Arg Arg Ser Gln Ser
                1505                1510                1515
Ser Leu Thr Ile


<210> 30
<211> 1585
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 6999183CD1

<400> 30
Met Ser Lys Arg Arg Met Ser Val Gly Gln Gln Thr Trp Ala Leu
 1               5                   10                  15
Leu Cys Lys Asn Cys Leu Lys Lys Trp Arg Met Lys Arg Gln Thr
                20                  25                  30
Leu Leu Glu Trp Leu Phe Ser Phe Leu Leu Val Leu Phe Leu Tyr
                35                  40                  45
Leu Phe Phe Ser Asn Leu His Gln Val His Asp Thr Pro Gln Met
                50                  55                  60
Ser Ser Met Asp Leu Gly Arg Val Asp Ser Phe Asn Asp Thr Asn
                65                  70                  75
Tyr Val Ile Ala Phe Ala Pro Glu Ser Lys Thr Thr Gln Glu Ile
                80                  85                  90
Met Asn Lys Val Ala Ser Ala Pro Phe Leu Met Ala Gly Arg Thr
                95                  100                 105
Ile Met Gly Trp Pro Asp Glu Lys Ser Met Asp Glu Leu Asp Leu
                110                 115                 120
Asn Tyr Ser Ile Asp Ala Val Arg Val Ile Phe Thr Asp Thr Phe
                125                 130                 135
Ser Tyr His Leu Lys Phe Ser Trp Gly His Arg Ile Pro Met Met
                140                 145                 150
Lys Glu His Arg Asp His Ser Ala His Cys Gln Ala Val Asn Glu
                155                 160                 165
Lys Met Lys Cys Glu Gly Ser Glu Phe Trp Glu Lys Gly Phe Val
                170                 175                 180
Ala Phe Gln Ala Ala Ile Asn Ala Ala Ile Ile Glu Ile Ala Thr
                185                 190                 195
Asn His Ser Val Met Glu Gln Leu Met Ser Val Thr Gly Val His
                200                 205                 210
Met Lys Ile Leu Pro Phe Val Ala Gln Gly Gly Val Ala Thr Asp
                215                 220                 225
Phe Phe Ile Phe Phe Cys Ile Ile Ser Phe Ser Thr Phe Ile Tyr
                230                 235                 240
Tyr Val Ser Val Asn Val Thr Gln Glu Arg Gln Tyr Ile Thr Ser
                245                 250                 255
Leu Met Thr Met Met Gly Leu Arg Glu Ser Ala Phe Trp Leu Ser
                260                 265                 270
```

Trp Gly Leu Met Tyr Ala Gly Phe Ile Leu Ile Met Ala Thr Leu
275 280 285
Met Ala Leu Ile Val Lys Ser Ala Gln Ile Val Val Leu Thr Gly
290 295 300
Phe Val Met Val Phe Thr Leu Phe Leu Leu Tyr Gly Leu Ser Leu
305 310 315
Ile Thr Leu Ala Phe Leu Met Ser Val Leu Ile Lys Lys Pro Phe
320 325 330
Leu Thr Gly Leu Val Val Phe Leu Leu Ile Val Phe Trp Gly Ile
335 340 345
Leu Gly Phe Pro Ala Leu Tyr Thr His Leu Pro Ala Phe Leu Glu
350 355 360
Trp Thr Leu Cys Leu Leu Ser Pro Phe Ala Phe Thr Val Gly Met
365 370 375
Ala Gln Leu Ile His Leu Asp Tyr Asp Val Asn Ser Asn Ala His
380 385 390
Leu Asp Ser Ser Gln Asn Pro Tyr Leu Ile Ile Ala Thr Leu Phe
395 400 405
Met Leu Val Phe Asp Thr Leu Leu Tyr Leu Val Leu Thr Leu Tyr
410 415 420
Phe Asp Lys Ile Leu Pro Ala Glu Tyr Gly His Arg Cys Ser Pro
425 430 435
Leu Phe Phe Leu Lys Ser Cys Phe Trp Phe Gln His Gly Arg Ala
440 445 450
Asn His Val Val Leu Glu Asn Glu Thr Asp Ser Asp Pro Thr Pro
455 460 465
Asn Asp Cys Phe Glu Pro Val Ser Pro Glu Phe Cys Gly Lys Glu
470 475 480
Ala Ile Arg Ile Lys Asn Leu Lys Lys Glu Tyr Ala Gly Lys Cys
485 490 495
Glu Arg Val Glu Ala Leu Lys Gly Val Val Phe Asp Ile Tyr Glu
500 505 510
Gly Gln Ile Thr Ala Leu Leu Gly His Ser Gly Ala Gly Lys Thr
515 520 525
Thr Leu Leu Asn Ile Leu Ser Gly Leu Ser Val Pro Thr Ser Gly
530 535 540
Ser Val Thr Val Tyr Asn His Thr Leu Ser Arg Met Ala Asp Ile
545 550 555
Glu Asn Ile Ser Lys Phe Thr Gly Phe Cys Pro Gln Ser Asn Val
560 565 570
Gln Phe Gly Phe Leu Thr Val Lys Glu Asn Leu Arg Leu Phe Ala
575 580 585
Lys Ile Lys Gly Ile Leu Pro His Glu Val Glu Lys Glu Val Leu
590 595 600
Leu Leu Asp Glu Pro Thr Ala Gly Leu Asp Pro Leu Ser Arg His
605 610 615
Arg Ile Trp Asn Leu Leu Lys Glu Gly Lys Ser Asp Arg Val Ile
620 625 630
Leu Phe Ser Thr Gln Phe Ile Asp Glu Ala Asp Ile Leu Ala Asp
635 640 645
Arg Lys Val Phe Ile Ser Asn Gly Lys Leu Lys Cys Ala Gly Ser
650 655 660
Ser Leu Phe Leu Lys Lys Lys Trp Gly Ile Gly Tyr His Leu Ser
665 670 675
Leu His Leu Asn Glu Arg Cys Asp Pro Glu Ser Ile Thr Ser Leu
680 685 690
Val Lys Gln His Ile Ser Asp Ala Lys Leu Thr Ala Gln Ser Glu
695 700 705
Glu Lys Leu Val Tyr Ile Leu Pro Leu Glu Arg Thr Asn Lys Phe
710 715 720
Pro Glu Leu Tyr Arg Asp Leu Asp Arg Cys Ser Asn Gln Gly Ile
725 730 735
Glu Asp Tyr Gly Val Ser Ile Thr Thr Leu Asn Glu Val Phe Leu

740 745 750

Lys Leu Glu Gly Lys Ser Thr Ile Asp Glu Ser Asp Ile Gly Ile

755 760 765

Trp Gly Gln Leu Gln Thr Asp Gly Ala Lys Asp Ile Gly Ser Leu

770 775 780

Val Glu Leu Glu Gln Val Leu Ser Ser Phe His Glu Thr Arg Lys

785 790 795

Thr Ile Ser Gly Val Ala Leu Trp Arg Gln Gln Val Cys Ala Ile

800 805 810

Ala Lys Val Arg Phe Leu Lys Leu Lys Lys Glu Arg Lys Ser Leu

815 820 825

Trp Thr Ile Leu Leu Leu Phe Gly Ile Ser Phe Ile Pro Gln Leu

830 835 840

Leu Glu His Leu Phe Tyr Glu Ser Tyr Gln Lys Ser Tyr Pro Trp

845 850 855

Glu Leu Ser Pro Asn Thr Tyr Phe Leu Ser Pro Gly Gln Gln Pro

860 865 870

Gln Asp Pro Leu Thr His Leu Leu Val Ile Asn Lys Thr Gly Ser

875 880 885

Thr Ile Asp Asn Phe Leu His Ser Leu Arg Arg Gln Asn Ile Ala

890 895 900

Ile Glu Val Asp Ala Phe Gly Thr Arg Asn Gly Thr Asp Asp Pro

905 910 915

Ser Tyr Asn Gly Ala Ile Ile Val Ser Gly Asp Glu Lys Asp His

920 925 930

Arg Phe Ser Ile Ala Cys Asn Thr Lys Arg Leu Asn Cys Phe Pro

935 940 945

Val Leu Leu Asp Val Ile Ser Asn Gly Leu Leu Gly Ile Phe Asn

950 955 960

Ser Ser Glu His Ile Gln Thr Asp Arg Ser Thr Phe Phe Glu Glu

965 970 975

His Met Asp Tyr Glu Tyr Gly Tyr Arg Ser Asn Thr Phe Phe Trp

980 985 990

Ile Pro Met Ala Ala Ser Phe Thr Pro Tyr Ile Ala Met Ser Ser

995 1000 1005

Ile Gly Asp Tyr Lys Lys Lys Ala His Ser Gln Leu Arg Ile Ser

1010 1015 1020

Gly Leu Tyr Pro Ser Ala Tyr Trp Phe Gly Gln Ala Leu Val Asp

1025 1030 1035

Val Ser Leu Tyr Phe Leu Ile Leu Leu Leu Met Gln Ile Met Asp

1040 1045 1050

Tyr Ile Phe Ser Pro Glu Glu Ile Ile Phe Ile Ile Gln Asn Leu

1055 1060 1065

Leu Ile Gln Ile Leu Cys Ser Ile Gly Tyr Val Ser Ser Leu Val

1070 1075 1080

Phe Leu Thr Tyr Val Ile Ser Phe Ile Phe Arg Asn Gly Arg Lys

1085 1090 1095

Asn Ser Gly Ile Trp Ser Phe Phe Phe Leu Ile Val Val Ile Phe

1100 1105 1110

Ser Ile Val Ala Thr Asp Leu Asn Glu Tyr Gly Phe Leu Gly Leu

1115 1120 1125

Phe Phe Gly Thr Met Leu Ile Pro Pro Phe Thr Leu Ile Gly Ser

1130 1135 1140

Leu Phe Ile Phe Ser Glu Ile Ser Pro Asp Ser Met Asp Tyr Leu

1145 1150 1155

Gly Ala Ser Glu Ser Glu Ile Val Tyr Leu Ala Leu Leu Ile Pro

1160 1165 1170

Tyr Leu His Phe Leu Ile Phe Leu Phe Ile Leu Arg Cys Leu Glu

1175 1180 1185

Met Asn Cys Arg Lys Lys Leu Met Arg Lys Asp Pro Val Phe Arg

1190 1195 1200

Ile Ser Pro Arg Ser Asn Ala Ile Phe Pro Asn Pro Glu Glu Pro

1205 1210 1215

```
Glu Gly Glu Glu Glu Asp Ile Gln Met Glu Arg Met Arg Thr Val
                1220                1225                1230
Asn Ala Met Ala Val Arg Asp Phe Asp Glu Thr Pro Val Ile Ile
                1235                1240                1245
Ala Ser Cys Leu Arg Lys Glu Tyr Ala Gly Lys Lys Lys Asn Cys
                1250                1255                1260
Phe Ser Lys Arg Lys Lys Thr Ile Ala Thr Arg Asn Val Ser Phe
                1265                1270                1275
Cys Val Lys Lys Gly Glu Val Ile Gly Leu Leu Gly His Asn Gly
                1280                1285                1290
Ala Gly Lys Ser Thr Thr Ile Lys Met Ile Thr Gly Asp Thr Lys
                1295                1300                1305
Pro Thr Ala Gly Gln Val Ile Leu Lys Gly Ser Gly Gly Gly Glu
                1310                1315                1320
Pro Leu Gly Phe Leu Gly Tyr Cys Pro Gln Glu Asn Ala Leu Trp
                1325                1330                1335
Pro Asn Leu Thr Val Arg Gln His Leu Glu Val Tyr Ala Ala Val
                1340                1345                1350
Lys Gly Leu Arg Lys Gly Asp Ala Met Ile Ala Ile Thr Arg Leu
                1355                1360                1365
Val Asp Ala Leu Lys Leu Gln Asp Gln Leu Lys Ala Pro Val Lys
                1370                1375                1380
Thr Leu Ser Glu Gly Ile Lys Arg Lys Leu Cys Phe Val Leu Ser
                1385                1390                1395
Ile Leu Gly Asn Pro Ser Val Val Leu Leu Asp Glu Pro Ser Thr
                1400                1405                1410
Gly Met Asp Pro Glu Gly Gln Gln Gln Met Trp Gln Val Ile Arg
                1415                1420                1425
Ala Thr Phe Arg Asn Thr Glu Arg Gly Ala Leu Leu Thr Thr His
                1430                1435                1440
Tyr Met Ala Glu Ala Glu Ala Val Cys Asp Arg Val Ala Ile Met
                1445                1450                1455
Val Ser Gly Arg Leu Arg Cys Ile Gly Ser Ile Gln His Leu Lys
                1460                1465                1470
Ser Lys Phe Gly Lys Asp Tyr Leu Leu Glu Met Lys Leu Lys Asn
                1475                1480                1485
Leu Ala Gln Met Glu Pro Leu His Ala Glu Ile Leu Arg Leu Phe
                1490                1495                1500
Pro Gln Ala Ala Gln Gln Glu Arg Phe Ser Ser Leu Met Val Tyr
                1505                1510                1515
Lys Leu Pro Val Glu Asp Val Arg Pro Leu Ser Gln Ala Phe Phe
                1520                1525                1530
Lys Leu Glu Ile Val Lys Gln Ser Phe Asp Leu Glu Glu Tyr Ser
                1535                1540                1545
Leu Ser Gln Ser Thr Leu Glu Gln Val Phe Leu Glu Leu Ser Lys
                1550                1555                1560
Glu Gln Glu Leu Gly Asp Leu Glu Glu Asp Phe Asp Pro Ser Val
                1565                1570                1575
Lys Trp Lys Leu Leu Leu Gln Glu Glu Pro
                1580                1585
```

<210> 31
<211> 1129
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2194064CB1

<400> 31
gcggccgcag ccttcgcgat aaacgggctg tcctacgggc tgctgcgctc gctgggcctt 60
gccttccctg accttgccga gcactttgac cgaagcgccc aggacactgc gtggatcagc 120

gccctggccc tggccgtgca gcaggcagcc agccccgtgg gcagcgccct gagcacgcgc 180
tggggggccc gccccgtggg tgatggttgg gggcgtcctc gcctcgctgg gcttcgtctt 240
ctcggctttc gccagcgatc tgctgcatct ctacctcggc ctgggcctcc tcgctggctt 300
tggttgggcc ctggtgttcg cccccgccct aggcaccctc tcgcgttact tctcccgccg 360
tcgagtcttg gcggtggggc tggcgctcac cggcaacggg gcctcctcgc tgctcctggc 420
gcccgccttg cagcttcttc tcgatacttt cggctggcgg ggcgctctgc tcctcctcgg 480
cgcgatcacc ctccacctca ccccctgtgg cgccctgctg ctacccctgg tccttcctgg 540
agaccccccca gccccaccgc gtagtcccct agctgccctc ggcctgagtc tgttcacacg 600
ccgggccttc tcaatctttg ctctaggcac agccctggtt gggggcgggt acttcgttcc 660
ttacgtgcac ttggctcccc acgctttaga ccggggcctg gggggatacg gagcagcgct 720
ggtggtggcc gtggctgcga tgggggatgc gggcgcccgg ctggtctgcg ggtggctggc 780
agaccaaggc tgggtgcccc tcccgcggct gctggccgta ttcggggctc tgactgggct 840
ggggctgtgg gtggtggggc tggtgcccgt ggtgggcggc gaagagagct ggggggggtcc 900
cctgctggcc gcggctgtgg cctatgggct gagcgcgggg agttacgccc cgctggtttt 960
cggtgtactc cccgggctgg tgggcgtcgg aggtgtggtg caggccacag ggctggtgat 1020
gatgctgatg agcctcgggg ggctcctggg ccctcccctg tcaggtaagg acctgagctc 1080
acagatctgc ctacaactat cctctgcccc tggggttcga ggcttctaa 1129

<210> 32
<211> 2699
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2744094CB1

<400> 32
agtgtgctgg aaagtgttat tattatttat aaaactgaac ctgcaggacc tgcataaggt 60
gaactaactt ctctagaaca ctgccctgtg ccaggcactg ggctgagtgc ttcacacaca 120
ttatatcata tcatcctcag catggctctg aagctggtaa ggttatcact atttcatagg 180
taaagaagtg gtatgttgct gacatttggt accttatcca aggtcacacg gtttggaagt 240
ggtggagcca ggatttgaac ccagggctct ctgactttgg aacccagctc ttcccattcc 300
tccaaggtgt ccacattggg tggggctcaa gggtttccaa gcactcaaag acagagatgg 360
ccgagcagct gagtcagcag ttgcctcgca cctgtttgtg gcacttgtac atcaccactg 420
tctccctccc aggttacatg gtctcttgca taattttctt ctttgtggtg ccgatcgtct 480
tcttaacgat cttcagcttc tggtggctga gctactggtt ggagcagggc tcggggacca 540
atagcagccg agagagcaat ggaaccatgg cagacctggg caacattgca gacaatcctc 600
aactgtcctt ctaccagctg gtgtacgggc tcaacgccct gctcctcatc tgtgtggggg 660
tctgctcctc agggattttc accaaggtca cgaggaaggc atccacggcc ctgcacaaca 720
agctcttcaa caaggttttc cgctgcccca tgagtttctt tgacaccatc ccaataggcc 780
ggcttttgaa ctgcttcgca ggggacttgg aacagctgga ccagctcttg cccatctttt 840
cagagcagtt cctggtcctg tccttaatgg tgatcgccgt cctgttgatt gtcagtgtgc 900
tgtctccata tatcctgtta atgggagcca taatcatggt tatttgcttc atttattata 960
tgatgttcaa gaaggccatc ggtgtgttca agagactgga gaactatagc cggtctcctt 1020
tattctccca catcctcaat tctctgcaag gcctgagctc catccatgtc tatggaaaaa 1080
ctgaagactt catcagccag tttaagaggc tgactgatgc gcagaataac tacctgctgt 1140
tgtttctatc ttccacacga tggatggcat tgaggctgga gatcatgacc aaccttgtga 1200
ccttggctgt tgccctgttc gtggcttttg gcatttcctc caccccctac tcctttaaag 1260
tcatggctgt caacatcgtg ctgcagctgg cgtccagctt ccaggccact gcccggattg 1320
gcttggagac agaggcacag ttcacggctg tagagaggat actgcagtac atgaagatgt 1380
gtgtctcgga agctccttta cacatggaag gcacaagttg tccccagggg tggccacagc 1440
atggggaaat catatttcag gattatcaca tgaaatacag agacaacaca cccaccgtgc 1500
ttcacggcat caacctgacc atccgcggcc acgaagtggt gggcatcgtg ggaaggacgg 1560
gctctgggaa gtcctccttg ggcatggctc tcttccgcct ggtggagccc atggcaggcc 1620
ggattctcat tgacggcgtg gacatttgca gcatcggcct ggaggacttg cggtccaagc 1680
tctcagtgat ccctcaagat ccagtgctgc tctcaggaac catcagattc aacctagatc 1740
cctttgaccg tcacactgac cagcagatct gggatgcctt ggagaggaca ttcctgacca 1800
aggccatctc aaagttcccc aaaaagctgc atacagatgt ggtggaaaac ggtggatact 1860
tctctgtggg ggagaggcag ctgctctgca ttgccagggc tgtgcttcgc aactccaaga 1920
tcatccttat cgatgaagcc acagcctcca ttgacatgga gacagacacc ctgatccagc 1980
gcacaatccg tgaagccttc cagggctgca ccgtgctcgt cattgcccac cgtgtcacca 2040
ctgtgctgaa ctgtgaccgc atcctggtta tgggcaatgg gaaggtggta gaatttgatc 2100

ggccggaggt actgcggaag aagcctgggt cattgttcgc agccctcatg gccacagcca 2160
cttcttcact gagataagga gatgtggaga cttcatggag gctggcagct gagctcagag 2220
gttcacacag gtgcagcttc gaggcccaca gtctgcgacc ttcttgtttg gagatgagaa 2280
cttctcctgg aagcaggggt aaatgtaggg ggggtgggga ttgctggatg gaaaccctgg 2340
aataggctac ttgatggctc tcaagacctt agaaccccag aaccatctaa gacatgggat 2400
tcagtgatca tgtggttctc cttttaactt acatgctgaa taattttata ataaggtaaa 2460
agcttatagt tttctgatct gtgttagaag tgttgcaaat gctgtactga ctttgtaaaa 2520
tataaaacta aggaaaactc actttctttg ttctgcttcc tttcgtttct ttcttttttg 2580
ttttttttaga cagggtcttg ctctgttgcc caggctggag cgcagtggcg caatctcagc 2640
tcactgtagc ctctgcctcc caggttcaag caattctcct gcctcagcct cctgaatag 2699

<210> 33
<211> 6369
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2798241CB1

<400> 33
tgccaccaca cccagctaat ttttatatgt ttagtagaga cagggtttca ccatgttggt 60
caggctggtc tcaaactcct gacttcgtga tctgcccacc ttggcctccc aaagtgctgg 120
gattacaggc gtgagccacc gcacccggtc agctattttc tacatgcttc atttgcagtg 180
taatattgga ttgtatgaga ctttgggttt tgtgttaata cctacagaaa atgttgatat 240
tttctcttag caggctgtca accaggttag gttcaggtca taagtttcta cccacattct 300
ttgaactgta gttgtcattt tagtttattt ttcaaaaact tttgcagtac ctttttggtc 360
tgtcttgtgt gtgccttgca gtgaacagtc tggatttgga cagtggtctg tctgttagtt 420
cagtttctca agcctttgtc acactaatag gattggattt atgtatgtcc agcttgggaa 480
ttattacagg aattaaaaac aactttttag agtgctttcc tgagctctct ttctatttgt 540
tccccccttct actttttgct tccctgtggc tgctgtttct atcctccagc cagagagcta 600
gtgtttattt tctccattgt gttacacact tgtgcagctg caaccaccat atccagggcc 660
caatggtagg aggtagagaa gaaaagcaaa agggattggc ctcatcctct tacaacgata 720
gttccattga atagagagaa aggtttccct gcctcagagt gttggctgca ctaggctttt 780
gttactgtag tctggccctg ttaccatggg attgcttgca tgtggggata caggagaatt 840
cagaaaagaa aaaaagattt gctatttcta cattctccct gagcattaag acttcccttg 900
cccattcctc aattcaaagc taaggcttct tctggagctg cctctgtggg cggttcggga 960
gataccaaag gagaaaaagt accactgttg atatggtggt atttcaaatt ctggtctacc 1020
ctatttcaca tgccttgttt acttttcaga gctgacagat tgctgctcca tgcattctgt 1080
ccagtttcct aagagagaca gcttggagta tgcttaatcc atcttacctg ggactgaaac 1140
agctgcttat tttgccgtta aaaattacat gcagtttact gcgtggctcc gggtttgttt 1200
gtttgttttt cctctttaat aggtttattc agaaaacatg tccactgcaa ttagggaggt 1260
aggagtttgg agacagacca gaacacttct actgaagaat tacttaatta aatgcagaac 1320
caaaaagagt agtgttcagg aaattctttt tccactattt tttttatttt ggttaatatt 1380
aattagcatg atgcatccaa ataagaaata tgaagaagtg cctaatatag aactcaatcc 1440
tatggacaag tttactcttt ctaatctaat tcttggatat actccagtga ctaatattac 1500
aagcagcatc atgcagaaag tgtctactga tcatctacct gatgtcataa ttactgaaga 1560
atatacaaat gaaaaagaaa tgttaacatc cagtctctct aagccgagca actttgtagg 1620
tgtggttttc aaagactcca tgtcctatga acttcgtttt tttcctgata tgattccagt 1680
atcttctatt tatatggatt caagagctgg ctgttcaaaa tcatgtgagg ctgctcagta 1740
ctggtcctca ggtttcacag ttttacaagc atccatagat gctgccatta tacagttgaa 1800
gaccaatgtt tctctttgga aggagctgga gtcaactaaa gctgttatta tgggagaaac 1860
tgctgttgta gaaatagata cctttccccg aggagtaatt ttaatatacc tagttatagc 1920
attttcacct tttggatact ttttggcaat tcatatcgta gcagaaaaag aaaaaaaaat 1980
aaaagaattt ttaaagataa tgggacttca tgatactgcc ttttggcttt cctgggttct 2040
tctatataca agtttaattt ttcttatgtc ccttcttatg gcagtcattg cgacagcttc 2100
tttgttattt cctcaaagta gcagcattgt gatatttctg ctttttttcc tttatggatt 2160
atcatctgta tttttttgctt taatgctgac acctcttttt aaaaaatcaa aacatgtggg 2220
aatagttgaa ttttttgtta ctgtggcttt tggatttatt ggccttatga taatcctcat 2280
agaaagtttt cccaaatcgt tagtgtggct tttcagtcct ttctgtcact gtacttttgt 2340
gattggtatt gcacaggtca tgcatttaga agattttaat gaaggtgctt cattttcaaa 2400
tttgactgca ggcccatatc ctctaattat tacaattatc atgctcacac ttaatagtat 2460
attctatgtc ctcttggctg tctatcttga tcaagtcatt ccaggggaat ttggcttacg 2520

gagatcatct ttatattttc tgaagccttc atattggtca aagagcaaaa gaaattatga 2580
ggagttatca gagggcaatg ttaatggaaa tattagtttt agtgaaatta ttgagccagt 2640
ttcttcagaa tttgtaggaa aagaagccat aagaattagt ggtattcaga agacatacag 2700
aaagaagggt gaaaatgtgg aggctttgag aaatttgtca tttgacatat atgagggtca 2760
gattactgcc ttacttggcc acagtggaac aggaaagagt acattgatga atattctttg 2820
tggactctgc ccaccttctg atgggtttgc atctatatat ggacacagag tctcagaaat 2880
agatgaaatg tttgaagcaa gaaaaatgat tggcatttgt ccacagttag atatacactt 2940
tgatgttttg acagtagaag aaaatttatc aattttggct tcaatcaaag ggataccagc 3000
caacaatata atacaagaag tgcagaaggt tttactagat ttagacatgc agactatcaa 3060
agataaccaa gctaaaaaat taagtggtgg tcaaaaaaga aagctgtcat taggaattgc 3120
tgttcttggg aacccaaaga tactgctgct agatgaacca acagctggaa tggacccctg 3180
ttctcgacat attgtatgga atcttttaaa atacagaaaa gccaatcggg tgacagtgtt 3240
cagtactcat ttcatggatg aagctgacat tcttgcagat aggaaagctg tgatatcaca 3300
aggaatgctg aaatgtgttg gttcttcaat gttcctcaaa agtaaatggg ggatcggcta 3360
ccgcctgagc atgtacatag acaaatattg tgccacagaa tctctttctt cactggttaa 3420
acaacatata cctggagcta ctttattaca acagaatgac caacaacttg tgtatagctt 3480
gcctttcaag gacatggaca aattttcagg tttgttttct gccctagaca gtcattcaaa 3540
tttgggtgtc atttcttatg gtgtttccat gacgactttg gaagacgtat ttttaaagct 3600
agaagttgaa gcagaaattg accaagcaga ttatagtgta tttactcagc agccactgga 3660
ggaagaaatg gattcaaaat cttttgatga aatggaacag agcttactta ttctttctga 3720
aaccaaggct tctctagtga gcaccatgag cctttggaaa caacagatgt atacaatagc 3780
aaagtttcat ttctttacct tgaaacgtga aagtaaatca gtgagatcag tgttgcttct 3840
gcttttaatt tttttcacag ttcagatttt tatgtttttg gttcatcact cttttaaaaa 3900
tgctgtggtt cccatcaaac ttgttccaga cttatatttt ctaaaacctg gagacaaacc 3960
acataaatac aaaacaagtc tgcttcttca aaattctgct gactcagata tcagtgatct 4020
tattagcttt ttcacaagcc agaacataat ggtgacgatg attaatgaca gtgactatgt 4080
atccgtggct ccccatagtg cggctttaaa tgtggtgcat tcagaaaagg actatgtttt 4140
tgcagctgtt ttcaacagta ctatggttta ttctttacct atattagtga atatcattag 4200
taactactat ctttatcatt taaatgtgac tgaaaccatc cagatctgga gtaccccatt 4260
ctttcaagaa attactgata tagtttttaa aattgagctg tattttcaag cagctttgct 4320
tggaatcatt gttactgcaa tgccacctta ctttgccatg gaaaatgcag agaatcataa 4380
gatcaaagct tatactcaac ttaaactttc aggtcttttg ccatctgcat attggattgg 4440
acaagctgtt gttgatatcc ccttattttt tatcattctt attttgatgc taggaagctt 4500
attggcattt cattatggat tatattttta tactgtaaag ttccttgctg tggttttttg 4560
ccttattggt tatgttccat cagttattct gttcacttat attgcttctt tcacctttaa 4620
gaaaatttta aataccaaag aattttggtc atttatctat tctgtggcag cgttggcttg 4680
tattgcaatc actgaaataa ctttctttat gggatacaca attgcaacta ttcttcatta 4740
tgccttttgt atcatcattc caatctatcc acttctaggt tgcctgattt ctttcataaa 4800
gatttcttgg aagaatgtac gaaaaaatgt ggacacctat aatccatggg ataggctttc 4860
agtagctgtt atatcgcctt acctgcagtg tgtactgtgg attttcctct tacaatacta 4920
tgagaaaaaa tatggaggca gatcaataag aaaagatccc tttttcagaa acctttcaac 4980
gaagtctaaa aataggaagc ttccagaacc accagacaat gaggatgaag atgaagatgt 5040
caaagctgaa agactaaagg tcaaagagct gatgggttgc cagtgttgtg aggagaaacc 5100
atccattatg gtcagcaatt tgcataaaga atatgatgac aagaaagatt ttcttctttc 5160
aagaaaagta aagaaagtgg caactaaata catctctttc tgtgtgaaaa aaggagagat 5220
cttaggacta ttgggtccaa atggtgctgg caaaagcaca attattaata ttctggttgg 5280
tgatattgaa ccaacttcag gccaggtatt tttaggagat tattcttcag agacaagtga 5340
agatgatgat tcactgaagt gtatgggtta ctgtcctcag ataaaccctt tgtggccaga 5400
tactacattg caggaacatt ttgaaattta tggagctgtc aaaggaatga gtgcaagtga 5460
catgaaagaa gtcataagtc gaataacaca tgcacttgat ttaaaagaac atcttcagaa 5520
gactgtaaag aaactacctg caggaatcaa acgaaagttg tgttttgctc taagtatgct 5580
agggaatcct cagattactt tgctagatga accatctaca ggtatggatc ccaaagccaa 5640
acagcacatg tggcgagcaa ttcgaactgc atttaaaaac agaaagcggg ctgctattct 5700
gaccactcac tatatggagg aggcagaggc tgtctgtgat cgagtagcta tcatggtgtc 5760
tgggcagtta agatgtatcg gaacagtaca acatctaaag agtaaatttg gaaaaggcta 5820
ctttttggaa attaaattga aggactggat agaaaaccta gaagtagacc gccttcaaag 5880
agaaattcag tatattttcc caaatgcaag ccgtcaggaa agtttttctt ctattttggc 5940
ttataaaatt cctaaggaag atgttcagtc cctttcacaa tcttttttta agctggaaga 6000
agctaaacat gcttttgcca ttgaagaata tagcttttct caagcaacat tggaacaggt 6060
ttttgtagaa ctcactaaag aacaagagga ggaagataat agttgtggaa ctttaaacag 6120
cacactttgg tgggaacgaa cacaagaaga tagagtagta ttttgaattt gtattgttcg 6180
gtctgcttac tgggacttct ttctttttca cttaatttta actttggttt aaaaagtttt 6240
ttattggaat ggtaactgga gaaccaagaa cgcacttgaa atttttctaa gctccttaat 6300

tgaaatgctg tggttgtgtg ttttgctttt ctttaaataa aacgtatgta taattaaaaa 6360
aaaaaaaaa 6369

<210> 34
<211> 2558
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 3105257CB1

<220>
<221> unsure
<222> 2523
<223> a, t, c, g, or other

<400> 34
atggggcgcg gggccggcgc tgctctgggg cgttggagcc gcgcgccgct ggaggagctg 60
ctgccggggc gggggtctgg gcggctcggg gggccacgcg ggcctcggac ggctcccggg 120
gctgtgggct tgggcccggc agctgcaggg gaggaggcct ggcggcgcgg gcgggcggcg 180
ccttcccggg acgaccagcg gctacgaccc atggcgcccg gactctcgga ggccgggaag 240
ctcctggggc tggagtaccc tgagcgccag aggctggcag ctgcggttgg atttctcacg 300
atgtccggtg ttatctccat gtctgcccct ttctttctgg ggaagatcat cgatgccatc 360
tataccaacc ccactgtgga ctacagcgac aacctgaccc gcctctgcct agggctcagt 420
gccgtgtttc tgtgtggtgc tgccgccaat gccattcgtg tctacctcat gcaaacttca 480
ggtcagcgca ttgtgaatag gctgagaact tcattattct cctccattct gaggcaggag 540
gttgctttct ttgacaagac tcgcacagga gaattgatta accgcctctc atcagacact 600
gcactcctgg ggcgctcagt gactgaaaac ctctcagatg ggctcagggc cggggcccag 660
gcttccgtag gcatcagtat gatgtttttt gtctcaccta atctggccac ctttgttttg 720
agcgtggtgc ctccagtgtc aatcattgct gtaatttatg ggcgatatct acggaaactg 780
accaaagtca ctcaggattc cctggcacaa gccactcagc tagctgagga acgtattgga 840
aatgtaagaa ctgttcgagc ttttgggaaa gaaatgactg aaatcgagaa atatgccagc 900
aaagtggacc atgtaatgca gttagcaagg aaagaggcat tcgcccgggc tggtttcttt 960
ggagcaactg ggctctccgg aaacctgatc gtgctttctg tcctgtacaa aggagggctg 1020
ctgatgggca gtgcccacat gaccgtgggt gaactctctt ccttcctaat gtatgctttc 1080
tgggttggaa taagcattgg aggtctgagc tctttctact cggagctgat gaaaggactg 1140
ggtgcagggg ggcgcctctg ggagctcctg gagagagagc ccaagctgcc ttttaacgag 1200
ggggtcatct taaatgagaa aagcttccag ggtgctttgg agtttaagaa cgtgcatttt 1260
gcctatccag ctcgcccaga ggtgcccata tttcaggatt tcagcctttc cattccgtca 1320
ggatctgtca cggcactggt tggcccaagt ggttctggca aatcaacagt gctttcactc 1380
ctgctgaggt tgtacgaccc tgcttctgga actattagtc ttgatggcca tgacatccgt 1440
cagctaaacc cagtgtggct gagatccaaa attgggacag tgagtcagga acccattttg 1500
ttttcttgct ctattgctga gaacattgct tatggtgctg atgacccttc ctctgtgacc 1560
gctgaggaaa tccagagagt ggctgaagtg gccaatacag tggccttcat ccggaatttc 1620
ccccaagggt tcaacactgt ggttggagaa aagggtgttc tcctctcagg tggcagaaa 1680
cagcggattg cgattgcccg tgctctgcta aagaatccca aaattcttct cctagatgaa 1740
gcaaccagtg cgctggatgc cgaaaatgag taccttgttc aagaagctct agatcgactg 1800
atggatggaa gaacggtgtt agttattgcc catcgtctgt ccaccattaa gaatgctaat 1860
atggttgctg ttcttgacca aggaaaaatt actgaatatg gaaaacatga agagctgctt 1920
tcaaaaccaa atgggatata cagaaaacta atgaacaaac aaagttttat ttcagcataa 1980
ggaagcaatt actggtaaac aatatgagac tttaatgcaa aacagtgttg cagaaaaaaa 2040
actcagagac tatgaaatac ataaaccata tatcaagtta tttgaaaaat acctattttt 2100
ttccaaagtg tgtaaaagat tgttttgaaa cgtacctgtt ctcaagatct ttttattcag 2160
agttttaata attgtaactt tttaaatgtc tatagcactg aagttatttt caggttttgt 2220
atttctttt cttgtggaat attttaatta atatagcatg gcacctcatt ttcttttgcc 2280
tgctgttaaa gattgaagct attgtcaaat gacaacttta aaaaggcaat tataaataaa 2340
aagcctgatt attttaggcc agtttaccaa tcactgtgta attcttctgg tagtattcta 2400
cctactttta agtctaattt taccgatcga ataacgcgct tgtgaatttt ataccttat 2460
tcggtaatct ctgaggaaac ctcttttttt acacccgcgg agaagggagt ttttttgccc 2520
ccncgggttt acagggggac acggaaatgt ccctcgaa 2558

<210> 35

<211> 5065
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 3200979CB1

<400> 35
atggttaaaa aagagataag cgtgcgtcaa caaattcagg ctcttctgta caagaatttt 60
cttaaaaaat ggagaataaa aagagagttt attgggctat atttgtgcat cttttcggaa 120
cacttcagag ctacccgttt tcctgaacaa cctcctaaag tcctgggaag cgtggatcag 180
tttaatgact ctggcctggt agtggcatat acaccagtca gtaacataac acaaaggata 240
atgaataaga tggccttggc ttcctttatg aaaggaagaa cagtcattgg gacaccagat 300
gaagagacca tggatataga acttccaaaa aaataccatg aaatggtggg agttatattt 360
agtgatactt tctcatatcg cctgaagttt aattggggat atagaatccc agttataaag 420
gagcactctg aatacacaga acactgttgg gccatgcatg gtgaaatttt ttgttacttg 480
gcaaagtact ggctaaaagg gtttgtagct tttcaagctg caattaatgc tgcaattata 540
gaagtcacaa caaatcattc tgtaatggag gagttgacat cagttattgg aataaatatg 600
aagataccac ctttcatttc taagggagaa attatgaatg aatggtttca ttttacttgc 660
ttagtttctt tctcttcttt tatatacttt gcatcattaa atgttgcaag ggaaagagga 720
aaatttaaga aactgatgac agtgatgggt ctccgagagt cagcattctg gctctcctgg 780
ggattgacat acatttgctt catcttcatt atgtccattt ttatggctct ggtcataaca 840
tcaatcccaa ttgtatttca tactggcttc atggtgatat tcacactcta tagcttatat 900
ggcctttctt tgatagcatt ggctttcctc atgagtgttt taataaggaa acctatgctc 960
gctggtttgg ctggatttct cttcactgta ttttggggat gtctgggatt cactgtgtta 1020
tatagacaac ttcctttatc tttgggatgg gtattaagtc ttcttagccc ttttgccttc 1080
actgctggaa tggcccagat tacacacctg gataattact taagtggtgt tatttttcct 1140
gatccctctg gggattcata caaaatgata gccacttttt tcattttggc atttgatact 1200
cttttctatt tgatattcac attatatttt gagcgagttt tacctgataa agatggccat 1260
ggggattctc cattattttt ccttaagtcc tcattttggt ccaaacatca aaatactcat 1320
catgaaatct ttgagaatga aataaatcct gagcattcct ctgatgattc ttttgaaccg 1380
gtgtctccag aattccatgg aaaagaagcc ataagaatca gaaatgttat aaaagaatat 1440
aatggaaaga ctggaaaagt agaagcattg caaggcatat tttttgacat atatgaagga 1500
cagatcactg caatacttgg gcataatgga gctggtaaat caacactgct aaacattctt 1560
agtggattgt ctgtttctac agaaggatca gccactattt ataatactca actctctgaa 1620
ataactgaca tggaagaaat tagaaagaat attggatttt gtccacagtt caattttcaa 1680
tttgacttcc tcactgtgag agaaaacctc agggtatttg ctaaaataaa agggattcag 1740
ccaaaggaag tggaacaaga ggttttgctg ctagatgaac caactgctgg attggatccc 1800
ttttcaagac accgagtgtg gagcctcctg aaggagcata aagtagaccg acttatcctc 1860
ttcagtaccc aattcatgga tgaggctgac atcttggctg ataggaaagt atttctgtct 1920
aatgggaagt tgaaatgtgc aggatcatct ttgtttctga agcgaaagtg gggtattgga 1980
tatcatttaa gtttacacag gaatgaaatg tgtgacacag aaaaaatcac atcccttatt 2040
aagcagcaca ttcctgatgc caagttaaca acagaaagtg aagaaaaact tgtatatagt 2100
ttgcctttgg aaaaaacgaa caaatttcca gatctttaca gtgaccttga taagtgttct 2160
gaccagggca taaggaatta tgctgtttca gtgacatctc tgaatgaagt attcttgaac 2220
ctagaaggaa aatcagcaat tgatgaacca gattttgaca ttgggaaaca agagaaaata 2280
catgtgacaa gaaatactgg agatgagtct gaaatggaac aggttctttg ttctcttcct 2340
gaaacaagaa aggctgtcag tagtgcagct ctctggagac gacaaatcta tgcagtggca 2400
acacttcgct tcttaaagtt aaggcgtgaa aggagagctc ttttgtgttt gttactagta 2460
cttggaattg cttttatccc catcattcta gagaagataa tgtataaagt aactcgtgaa 2520
actcattgtt gggagttttc acccagtatg tatttccttt ctctggaaca aatcccgaag 2580
acgcctctta ccagcctgtt aatcgttaat aatacaggat caaatattga agacctcgtg 2640
cattcactga agtgtcagga tatagttttg gaaatagatg actttagaaa cagaaatggc 2700
tcagatgatc cctcctacaa tggagccatc atagtgtctg gtgaccagaa ggattacaga 2760
ttttcagttg catgtaatac caagaaatcg aattgttttc ctgttcttat gggaattgtt 2820
agcaatgccc ttattggaat ttttaacttc acagagctta ttcaaatgga gagcacttca 2880
ttttttcgtg atgacatagt gctggatctt ggttttatag atgggtccat atttttgttg 2940
ttgatcacaa actgcatttc tccttatatt ggcataagca gcatcagtga ttataaaatc 3000
ccttcctcta tcccttctat tctttgtcag aaaaatgttc aatcccagtt atggatttca 3060
ggcctctggc cttcagcata ctggtgtgga caggctctgg tggacattcc attacacttc 3120
ttgattctcc tttcaataca tttaatttac tacttctcat ttctgggatt ccagcttcca 3180
tgggaactca tgtttgtttt ggtggtatgc ataattggtt gtgcagcttc tcttatattc 3240

ctcatgtacg tgctttcatt catcttttgc aagtggagaa aaaataatgg cttttggtct 3300
tttggctttt ttattgtctt aatatgtgta tccacaattc tggtatcaac taagtatgaa 3360
aaacccaact taattttgtg catgatattc ataccttcct ttactttcct agatatgtcg 3420
ttattgatcc agctcaactt tatgtatatg agaaacttgg acagtctgga caatagaata 3480
aatgaagtca ataaaaccat tcttttaaca aacttaatac cataccttca gagtgttatt 3540
ttccttttg tcataaggtg tctggaaatg aagtatggaa atgaaatcat gaataaagac 3600
ccagttttca gaatctctcc acgaagtaga ggaactcata ccaatccaga agagcctgaa 3660
gaagatgttc aagctgaaag agtccaagca gcaaatgcac tcactactcc aaacttggag 3720
gaggaaccag tcataactgc aagctgttta cacaaggaat attatgagac aaagaaaagt 3780
tgcttttcaa caacaaagaa gaaagcagcc atcagaaatg tttcgttttg tgttaaaaaa 3840
ggtgaagttt tgggattact aggacacaat ggagctggca aaagtacttc cattaaaatg 3900
ataactgggt gcacagtgcc aactgcagga gtggtggtgt tacaaggcaa cagagcatca 3960
gtaaggcaac agcgtgacaa cagcctcaag ttcttggggt actgccctca ggagaactca 4020
ctgtggccca agcttacaat gaaagagcac ttggagttgt atgcagctgt gaaaggactg 4080
ggcaaagaag atgctgctct cagtatttca cgattggtgg aagctcttaa gctccaggaa 4140
caacttaagg ctcctgtgaa aactctatca gagggaataa agagaaagct gtgctttgtg 4200
ctgagcatcc tggggaaccc atcagtggtg cttctagatg agccgttcac cgggatggac 4260
cccgaggggc agcagcaaat gtggcagata cttcaggcta ccattaaaaa ccaggagagg 4320
ggcaccctct tgaccaccca ttacatgtca gaggctaagt ctctgtgtga ccgtgtggcc 4380
atcatggtgt caggaacgct aaggtgtatt ggttccattc aacatctgaa aaacaagttt 4440
ggtaaagatt atttactaga aataaaaatg aaagaaccta ctcaggtgga agctctccac 4500
acagagattt tgaagctttt cccacaggct gcttggcagg aaagatattc ctctttaatg 4560
gcgtataagt tacctgtgga ggatgtccac cctctatctc gggccttttt caagttagag 4620
gcgatgaaac agaccttcaa cctggaggaa tacagcctct ctcaggctac cttggagcag 4680
gtattcttag aactctgtaa agagcaggag ctgggaaatg ttgatgataa aattgataca 4740
acagttgaat ggaaacttct cccacaggaa gacccttaaa atgaagaacc tcctaacatt 4800
caattttagg tcctactaca ttgttagttt ccataattct acaagaatgt ttcctttac 4860
ttcagttaac aaaagaaaat attcaatagt ttaaacatgc aacaatgatt acagttttca 4920
tttttaaaaa tttaggatga aggaaaaagg aaatataggg aaaagtagta gacaaaatta 4980
acaaaatcag acatgttatt atccccaaca tgggtctatt ttgtgcttag gggatccact 5040
agtttagaac gccggaccgc gtggt 5065

<210> 36
<211> 1677
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 6754139CB1

<400> 36
gtggaaacag gtttgagaga tactggaggg ggcagagcag tggggtttag aatccctggg 60
tgaaagtctg gactcttgtg gcttatttgg gcccctctag catttgtgga gaggcaggca 120
gactccaggt ccttgaaaag gggagggtgg aggagaaatt tgtcagcctg gcgccagaag 180
atagtaccag ttcactccat ggcctttacc tcatgtgtcc ctgcaggcag gccagggagg 240
aactagagcc acagctagag caagagaagg cagacaccag gaggacactc ataaggacag 300
ggccccagcc ctgggagtgg agggtgtgag cagaggccct gggactaggg cctgggatgg 360
acaaccctcc ttactgaccc tccagagtgc ctgggagctg agggccggct ggctctcaag 420
ctgttccgtg acctctttgc caactacaca agtgccctga gacctgtggc agacacagac 480
cagactctga atgtgaccct ggaggtgaca ctgtcccaga tcatcgacat ggatgaacgg 540
aaccaggtgc tgaccctgta tctgtggata cggcaggagt ggacagatgc ctacctacga 600
tgggacccca atgcctatgg tggcctggat gccatccgca tccccagcag tcttgtgtgg 660
cggccagaca tcgtactcta taacaaagcc gacgcgcagc ctccaggttc cgccagcacc 720
aacgtggtcc tgcgccacga tggcgccgtg cgctgggacg cgccggccat cacgcgcagc 780
tcgtgccgcg tggatgtagc agccttcccg ttcgacgccc agcactgcgg cctgacgttc 840
ggctcctgga ctcacggcgg gcaccaagtg gatgtgcggc cgcgcggcgc tgcagccagc 900
ctggcggact tcgtggagaa cgtggagtgg cgcgtgctgg gcatgccggc gcggcggcgc 960
gtgctcacct acggctgctg ctccgagccc taccccgacg tcaccttcac gctgctgctg 1020
cgccgccgcg ccgccgccta cgtgtgcaac ctgctgctgc cctgcgtgct catctcgctg 1080
cttgcgccgc tcgccttcca cctgcctgcc gactcaggcg agaaggtgtc gctgggcgtc 1140
accgtgctgc tggcgctcac cgtcttccag ttgctgctgg ccgagagcat gccaccggcc 1200
gagagcgtgc cgctcatcgg gaagtactac atggccacta tgaccatggt cacattctca 1260

acagcactca ccatccttat catgaacctg cattactgtg gtcccagtgt ccgcccagtg 1320
ccagcctggg ctagggccct cctgctggga cacctggcac ggggcctgtg cgtgcgggaa 1380
agaggggagc cctgtgggca gtccaggcca cctgagttat ctcctagccc ccagtcgcct 1440
gaaggagggg ctggcccccc agcgggccct tgccacgagc cacgatgtct gtgccgccag 1500
gaagccctac tgcaccacgt agccaccatt gccaatacct tccgcagcca ccgagctgcc 1560
cagcgctgcc atgaggactg gaagcgcctg gcccgtgtga tggaccgctt cttcctggcc 1620
atcttcttct ccatggccct ggtcatgagc ctcctggtgc tggtgcaggc cctgtga 1677

<210> 37
<211> 3714
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 6996659CB1

<400> 37
atgaggagac tgagtttgtg gtggctgctg agcagggtct gtctgctgtt gccgccgccc 60
tgcgcactgg tgctggccgg ggtgcccagc tcctcctcgc acccgcagcc ctgccagatc 120
ctcaagcgca tcgggcacgc ggtgagggtg ggcgcggtgc acttgcagcc ctggaccacc 180
gccccccgcg cggccagccg cgctccggac gacagccgag caggagccca gagggatgag 240
ccggagccag ggactaggcg gtccccggcg ccctcgccgg gcgcacgctg gttggggagc 300
accctgcatg gccggggggcc gccgggctcc cgtaagcccg gggagggcgc cagggcggag 360
gccctgtggc cacgggacgc cctcctattt gccgtggaca acctgaaccg cgtggaaggg 420
ctgctaccct acaacctgtc tttggaagta gtgatggcca tcgaggcagg cctgggcgat 480
ctgccacttt tgcccttctc ctcccctagt tcgccatgga gcagtgaccc tttctccttc 540
ctgcaaagtg tgtgccatac cgtggtggtg caaggggtgt cggcgctgct cgccttcccc 600
cagagccagg gcgaaatgat ggagctcgac ttggtcagct tagtcctgca cattccagtg 660
atcagcatcg tgcgccacga gtttccgcgg gagagtcaga atccccttca cctacaactg 720
agtttagaaa attcattaag ttctgatgct gatgtcactg tctcaatcct gaccatgaac 780
aactggtaca attttagctt gttgctgtgc caggaagact ggaacatcac cgatttcctc 840
ctccttaccc agaataattc caagttccac cttggttcta tcatcaacat caccgctaac 900
ctcccctcca cccaggacct cttgagcttc ctacagatcc agcttgagag tattaagaac 960
agcacaccca cagtggtgat gtttggctgc gacatggaaa gtatccggcg gattttcgaa 1020
attacaaccc agtttggggt catgcccccct gaacttcgtt gggtgctggg agattcccag 1080
aatgtggagg aactgaggac agagggtctg cccttaggac tcattgctca tggaaaaaca 1140
acacagtctg tctttgagca ctacgtacaa gatgctatgg agctggtcgc aagagctgta 1200
gccacagcca ccatgatcca accagaactt gctctcattc ccagcacgat gaactgcatg 1260
gaggtggaaa ctacaaatct cacttcagga caatatttat caaggtttct agccaatacc 1320
actttcagag gcctcagtgg ttccatcaga gtaaaaggtt ccaccatcgt cagctcagaa 1380
aacaactttt tcatctggaa tcttcaacat gaccccatgg gaaagccaat gtggacccgc 1440
ttgggcagct ggcagggggg aaagattgtc atggactatg gaatatggcc agagcaggcc 1500
cagagacaca aaacccactt ccaacatcca agtaagctac acttgagagt ggttaccctg 1560
attgagcatc cttttgtctt cacaagggag gtagatgatg aaggcttgtg ccctgctggc 1620
caactctgtc tagaccccat gactaatgac tcttccacat tggacagcct tttttagcagc 1680
ctccatagca gtaatgatac agtgcccatt aaattcaaga agtgctgcta tggatattgc 1740
attgatctgc tggaaaagat agcagaagac atgaactttg acttcgacct ctatattgta 1800
ggggatggaa agtatggagc atggaaaaat gggcactgga ctgggctagt gggtgatctc 1860
ctgagaggga ctgcccacat ggcagtcact tcctttagca tcaatactgc acggagccag 1920
gtgatagatt tcaccagccc tttcttctcc accagcttgg gcatcttagt gaggacccga 1980
gatacagcag ctcccattgg agccttcatg tggccactcc actggacaat gtggctgggg 2040
atttttgtgg ctctgcacat cactgccgtc ttcctcactc tgtatgaatg gaagagtcca 2100
tttggtttga cttccaaggg gcgaaataga agtaaagtct tctccttttc ttcagccttg 2160
aacatctgtt atgccctctt gtttggcaga acagtggcca tcaaacctcc aaaatgttgg 2220
actggaaggt ttctaatgaa cctttgggcc attttctgta tgttttgcct ttccacatac 2280
acggcaaact tggctgctgt catggtaggt gagaagatct atgaagagct ttctggaata 2340
catgacccca agttacatca tccttcccaa ggattccgct ttggaactgt ccgagaaagc 2400
agtgctgaag attatgtgag acaaagtttc ccagagatgc atgaatatat gagaaggtac 2460
aatgttccag ccacccctga tggagtggag tatctgaaga atgatccaga gaaactagac 2520
gccttcatca tggacaaagc ccttctggat tatgaagtgt caatagatgc tgactgcaaa 2580
cttctcactg tggggaagcc atttgccata gaaggttacg gcattggcct cccacccaac 2640
tctccattga ccgccaacat atccgagcta atcagtcaat acaagtcaca tgggtttatg 2700

gatatgctcc atgacaagtg gtacagggtg gttccctgtg gcaagagaag ttttgctgtc 2760
acggagactt tgcaaatggg catcaaacac ttctctgggc tctttgtgct gctgtgcatt 2820
ggatttggtc tgtccatttt gaccaccatt ggtgagcaca tagtatacag gctgctgcta 2880
ccacgaatca aaaacaaatc caagctgcaa tactggctcc acaccagcca gagattacac 2940
agagcaataa atacatcatt tatagaggaa aagcagcagc atttcaagac caaacgtgtg 3000
gaaaagaggt ctaatgtggg accccgtcag cttaccgtat ggaatacttc caatctgagt 3060
catgacaacc gacggaaata catctttagt gatgaggaag gacaaaacca gctaggcatc 3120
cggatccacc aggacatccc cctccctcca aggagaagag agctccctgc cttgcggacc 3180
accaatggga aagcagactc cctaaatgta tctcggaact cagtgatgca ggaactctca 3240
gagctcgaga agcagattca ggtgatccgt caggagctgc agctggctgt gagcaggaaa 3300
acggagctgg aggagtatca aaggacaagt cggacttgtg agtcctaggt gaccacactg 3360
cttccctttc tcagttcctg accttcctct gagcccttga gacactttgt aatgctcttt 3420
tgtaactatc gacaaaggtg tggggaagct gaggtctagg tcttcttaaa ggtcaagtct 3480
gctctccctc gcctaaagtg cagcagcagc tcctctcaag ctcactctct aggtctccag 3540
ggtaggagtg tttttctagc aagaatctta gtcaggagta agctctgtgc gagagatctg 3600
tgaataacca gataacccca gctgccgtta accttttcac caggtgccac agtaatattt 3660
ctggttttta gccctttctc tgcactacca acaagagata aaattgttac tcac 3714

<210> 38
<211> 1009
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7472747CB1

<400> 38
cgaggaggag ggcagggcag gcggcagcct gggcacaggc ccctaggtgc ttactcctca 60
cctgtttccc acctctcccc catagccagc cccacggccc tggcagggtc ctggccacag 120
catgcccagt gctgggctct gcagctgctg gggtggccgg gtgctgcccc tgctgctggc 180
ctatgtctgc tacctgctgc tcggtgccac tatcttccag ctgctagaga ggcaggcgga 240
ggctcagtcc agggaccagt ttcagttgga gaagctgcgc ttcctggaga actacacctg 300
cctggaccag tgggccatgg agcagtttgt gcaggtcatc atggaagcct gggtgaaagg 360
tgtgaacccc aaaggcaact ctaccaaccc cagcaactgg gactttggca gcagtttctt 420
ctttgcaggc acagtcgtca ctaccatagg ttatgggaac ctggcaccca gcacagaggc 480
aggtcaggtc ttctgtgtct tctatgccct gttgggcatc ccgcttaacg tgatcttcct 540
caaccacctg ggcacagggc tgcgtgccca tctggccgcc attgaaagat gggaggaccg 600
tcccaggcgc tcccaggagg tactgcaagt cctgggcctg gctctgttcc tgaccctggg 660
gacgctggtc attctcatct tcccacccat ggtcttcagc catgtggagg gctggagctt 720
cagcgagggc ttctactttg ctttcatcac tctcagcacc attggctttg ggggactatgt 780
tgcaggcaca gaccccagca agcattatat ctcagtgtat cggagcctgg cagccatctg 840
gatcctcctg ggcctggcgt ggctggcgct gatcctccca ctgggccccc tgcttctgca 900
cagatgctgc cagctctggc tgctcagtag ggcctcggc gtcaaggatg ggcagcctc 960
tgaccccagt gggctcccca ggcctcagaa gatccccatc tctgcatga 1009

<210> 39
<211> 1155
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7474121CB1

<400> 39
atggaggtct cggggcaccc ccaggccagg agatgctgcc cagaggccct gggaaagctc 60
ttccctggcc tctgcttcct ctgctttctg gtgacctacg ccctggtggg tgctgtggtc 120
ttctctgcca ttgaggacgg ccaggtcctg gtggcagcag atgatggaga gtttgagaag 180
ttcttggagg agctctgcag aatcttgaac tgcagtgaaa cagtggtgga agacagaaaa 240
caggatctcc aggggcatct gcagaaggtg aagcctcagt ggtttaacag gaccacacac 300
tggtccttcc tgagctcgct ctttttctgc tgcacggtgt tcagcaccgt gggctatggc 360
tacatctacc ccgtcaccag gcttggcaag tacttgtgca tgctctatgc tctctttggt 420

atccccctga tgttcctcgt tctcacggac acaggcgaca tcctggcaac catcttatct 480
acatcttata atcggttccg aaaattccct ttctttaccc gccccctcct ctccaagtgg 540
tgccccaaat ctctcttcaa gaaaaaaccg gaccccaagc ccgcagatga agctgtccct 600
cagatcatca tcagtgctga agagcttcca ggccccaaac ttggcacatg tccttcacgc 660
ccaagctgca gcatggagct gtttgagaga tctcatgcgc tagagaaaca gaacacactg 720
caactgcccc cacaagccat ggagaggagt aactcgtgtc ccgaactggt gttgggaaga 780
ctctcatact ccatcatcag caacctggat gaagttggac agcaggtgga gaggttggac 840
atccccctcc ccatcattgc ccttattgtt tttgcctaca tttcctgtgc agctgccatc 900
ctcccccttct gggagacaca gttggatttc gagaatgcct tctatttctg ctttgtcaca 960
ctcaccacca ttgggtttgg ggatactgtt ttagaacacc ctaacttctt cctgttcttc 1020
tccatttata tcatcgttgg aatggagatt gtgttcattg ctttcaagtt ggtgcaaaac 1080
aggctgattg acatatacaa aaatgttatg ctattctttg caaaagggaa gttttaccac 1140
cttgttaaaa agtga 1155

<210> 40
<211> 2733
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7475615CB1

<400> 40
cccttcattg agctccttca ccagcaccag gaaggcaccg ctgatgagag cgaagatgag 60
cgaggcgatg ttggtgtggg ggaggttttt gcaaatgtca atgaaggtaa agacgatgga 120
ccctgggcct gtgtaggagg ggatggtcag tccgaagatg tacttgagca ccgaaatcag 180
gaacaccttt cggctgcccg ctccccagac acatctgcag ccctgcccag ccggctttgc 240
tcacccactg cttgtaaatg ccccagatat gagccagccc aggccccgct acgtggtaga 300
cagagccgca tactccctta ccctcttcga cgatgagttt gagaagaagg accggacata 360
cccagtggga gagaaacttc gcaatgcctt cagatgttcc tcagccaaga tcaaagctgt 420
ggtgtttggg ctgctgcctg tgctctcctg gctccccaag tacaagatta aagactacat 480
cattcctgac ctgctcggtg gactcagcgg gggatccatc caggtcccac aaggcatggc 540
atttgctctg ctggccaacc ttcctgcagt caatggcctc tactcctcct tcttccccct 600
cctgacctac ttcttcctgg ggggtgttca ccagatggtg ccaggtacct ttgccgttat 660
cagcatcctg gtgggtaaca tctgtctgca gctggcccca gagtcgaaat tccaggtctt 720
caacaatgcc accaatgaga gctatgtgga cacagcagcc atggaggctg agaggctgca 780
cgtgtcagct acgctagcct gcctcaccgc catcatccag atgggtctgg gcttcatgca 840
gtttggcttt gtggccatct acctctccga gtccttcatc cggggcttca tgacggccgc 900
cggcctgcag atcctgattt cggtgctcaa gtacatcttc ggactgacca tcccctccta 960
cacaggccca gggtccatcg tctttacctt cattgacatt tgcaaaaacc tcccccacac 1020
caacatcgcc tcgctcatct tcgctctcat cagcggtgcc ttcctggtgc tggtgaagga 1080
gctcaatgct cgctacatgc acaagattcg cttccccatc cctacagaga tgattgtggt 1140
ggtggtggca acagctatct ccgggggctg taagatgccc aaaaagtatc acatgcagat 1200
cgtgggagaa atccaacgcg ggttccccac cccggtgtcg cctgtggtct cacagtggaa 1260
ggacatgata ggcacagcct tctccctagc catcgtgagc tacgtcatca acctggctat 1320
gggccggacc ctggccaaca agcacggcta cgacgtggat tcgaaccagg agatgatcgc 1380
tctcggctgc agcaacttct ttggctcctt ctttaaaatt catgtcattt gctgtgcgct 1440
ttctgtcact ctggctgtgg atggagctgg aggaaaatcc cagtctgtgc taggagccct 1500
gatcgctgtc aatctcaaga actccctcaa gcaactcacc gacccctact acctgtggag 1560
gaagagcaag ctggactgtt gcatctgggt agtgagcttc ctctcctcct tcttcctcag 1620
cctgccctat ggtgtggcag tgggtgtcgc cttctccgtc ctggtcgtgg tcttccagac 1680
tcagtttcga aatggctatg cactggccca ggtcatggac actgacattt atgtgaatcc 1740
caagacctat aatagggccc aggatatcca ggggattaaa atcatcacgt actgctcccc 1800
tctctacttt gccaactcag agatcttcag gcaaaaggtc atcgccaaga ctgtctccct 1860
gcaggagctg cagcaggact ttgagaatgc gccccccacc gaccccaaca acaaccagac 1920
cccggctaac ggcaccagcg tgtcctatat caccttcagc cctgacagct cctcacctgc 1980
ccagagtgag ccaccagcct ccgctgaggc cccggcgag cccagtgaca tgctggccag 2040
cgtcccaccc ttcgtcacct tccacaccct catcctggac atgagtggag tcagcttcgt 2100
ggacttgatg ggcatcaagg ccctggccaa gctgagctcc acctatggga agatcggcgt 2160
gaaggtcttc ttggtgaaca tccatgccca ggtgtacaat gacattagcc atggaggcgt 2220
ctttgaggat gggagtctag aatgcaagca cgtctttccc agcatacatg acgcagtcct 2280
ctttgcccag gcaaatgcta gagacgtgac cccaggacac aacttccaag ggctccagg 2340

ggatgctgag ctctccttgt acgactcaga ggaggacatt cgcagctact gggacttaga 2400
gcaggagatg ttcgggagca tgtttcacgc agagaccctg accgccctgt gagggctcag 2460
ccagtcctca tgctgcctac agagtgcctg gcacttggga cttccataaa ggatgagcct 2520
ggggtcacag ggggtgtcgg gcggaggaaa gtgcatcccc cagagcttgg gttcctctct 2580
cctctccccc tctctcctcc cttccttccc tccccgcatc tccagagaga gcctctcagc 2640
agcagggggg tgctaccctt acaggagtga gagtctggtg agcccactct tcacccgtca 2700
ggcctggccg caatggacaa gcctcctgct cac 2733

<210> 41
<211> 3457
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7475656CB1

<400> 41
cgagtctgga gcccgcgccg tcgccggccg cgtcctccgg gcatggaagg aggcggcaag 60
cccaactctt cgtctaacag ccgggacgat ggcaacagcg tcttccccgc caaggcgtcc 120
gcgccgggcg cggggccggc cgcggccgag aagcgcctgg gcaccccgcc gggggggcggc 180
ggggccggcg cgaaggagca cggcaactcc gtgtgcttca aggtggacgg cggtggcggc 240
gaggagccgg cggggggctt cgaagacgcc gaggggcccc ggcggcagta cggcttcatg 300
cagaggcagt tcacctccat gctgcagccc ggggtcaaca aattctccct ccgcatgttt 360
gggagccaga aggcggtgga aaaggagcag gaaagggtta aaactgcagg cttctggatt 420
atccaccctt acagtgattt caggttttac tgggatttaa taatgcttat aatgatggtt 480
ggaaatctag tcatcatacc agttggaatc acattcttta cagagcaaac aacaacacca 540
tggattattt tcaatgtggc atcagataca gttttcctat tggacctgat catgaatttt 600
aggactggga ctgtcaatga agacagttct gaaatcatcc tggaccccaa agtgatcaag 660
atgaattatt taaaaagctg gtttgtggtt gacttcatct catccatccc agtggattat 720
atctttctta ttgtagaaaa aggaatggat tctgaagttt acaagacagc cagggcactt 780
cgcattgtga ggtttacaaa aattctcagt ctcttgcgtt tattacgact ttcaaggtta 840
attagataca tacatcaatg ggaagagata ttccacatga catatgatct cgccagtgca 900
gtggtgagaa tttttaatct catcggcatg atgctgctcc tgtgccactg ggatggttgt 960
cttcagttct tagtaccact actgcaggac ttcccaccag attgctgggt gtctttaaat 1020
gaaatggtta atgattcttg ggaaagcag tattcatacg cactcttcaa agctatgagt 1080
cacatgctgt gcattgggta tggagcccaa gccccagtca gcatgtctga cctctggatt 1140
accatgctga gcatgatcgt cggggccacc tgctatgcca tgtttgtcgg ccatgccacc 1200
gctttaatcc agtctctgga ttcttcgagg cggcagtatc aagagaagta taagcaagtg 1260
gaacaataca tgtcattcca taagttacca gctgatatgc gtcagaagat acatgattac 1320
tatgaacaca gataccaagg caaaatcttt gatgaggaaa atattctcaa tgaactcaat 1380
gatcctctga gagaggagat agtcaacttc aactgtcgga aactggtggc tacaatgcct 1440
ttatttgcta atgcggatcc taattttgtg actgccatgc tgagcaagtt gagatttgag 1500
gtgtttcaac ctggagatta tatcatacga gaaggagccg tgggtaaaaa aatgtatttc 1560
attcaacacg gtgttgctgg tgtcattaca aaatccagta aagaaatgaa gctgacagat 1620
ggctcttact ttggagagat ttgcctgctg accaaaggac gtcgtactgc cagtgttcga 1680
gctgatacat attgtcgtct ttactcactt tccgtggaca atttcaacga ggtcctggag 1740
gaatatccaa tgatgaggag agcctttgag acagttgcca ttgaccgact agatcgaata 1800
ggaaagaaaa attcaattct tctgcaaaag ttccagaagg atctgaacac tggtgttttc 1860
aacaatcagg agaacgaaat cctcaagcag attgtgaaac atgacaggga gatggtgcag 1920
gcaatcgctc ccatcaatta tcctcaaatg acaaccctga attccacatc gtctactacg 1980
accccgacct cccgcatgag gacacaatct ccaccggtgt acacagcgac cagcctgtct 2040
cacagcaacc tgcactcccc cagtcccagc acacagaccc cccagccatc agccatcctg 2100
tcaccctgct cctacaccac cgcggtctgc agccctcctg tacagagccc tctggccgct 2160
cgaactttcc actatgcctc ccccaccgcc tcccagctgt cactcatgca acagcagccg 2220
cagcagcagg tacagcagtc ccagccgccg cagactcagc cacagcagcc gtccccgcag 2280
ccacagacac ctggcagctc cacgccgaaa aatgaagtgc acaagagcac gcaggcgctt 2340
cacaacacca acctgacccg ggaagtcagg ccactctccg cctcgcagcc ctcgctgccc 2400
catgaggtgt ccactctgat ttccagacct catcccactg tgggcgagtc cctggcctcc 2460
atccctcaac ccgtgacggc ggtccccgga acgggccttc aggcaggggg caggagcact 2520
gtcccgcagc gcgtcaccct cttccgacag atgtcgtcgg gagccatccc cccgaaccga 2580
ggagtccctc cagcaccccc tccaccagca gctgctcttc caagagaatc ttcctcagtc 2640
ttaaacacag acccagacgc agaaaagcca cgatttgctt caaatttatg atccctgctg 2700

attgtcaaag cagaaagaaa tactctcata aactgagact atactcagat cttattttat 2760
tctatctcct gatagatccc tctagcctac tatgaagaga tattttagac agctgtggcc 2820
tacacgtgaa atgtaaaaat atatatacat atactataaa atatatatct aaattcccaa 2880
gagagggtca aaagacctgt ttagcattca gtgttatatg tcttcctttc tttaaatcat 2940
taaaggattt aaaatgtcgt tgtaagatta tttatttcta acctactttt acttaagtcc 3000
tttgatatgt atatttctct attttatgaa gagttcttgg attcaatgga aacaaaactg 3060
attttaaaaa ggcaactcaa atgaactagt aaatagcacc aatcaaaact ttctttcatt 3120
agctgtgtct ctgcatctaa attgttaatc attaatggtg gagaattaaa taacaaatcc 3180
cattttatag atctaaattg tatttcggtg ctttcaattt caaattaggt taaagaatgc 3240
actacttgct tggccaccgt aggagactag cattgccact gtttgttaag aatatcacta 3300
acctcaaaca tgttcattga tctttcagaa agctgaggga aaattaatat ttgtcttcat 3360
gtgttatcgg acttttacca agactcgatc aatgttagtt gtaaataact ttttcaaccc 3420
aaataaaaat agctattctg tgttgtaaaa aaaaaaa 3457

<210> 42
<211> 5622
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7480632CB1

<400> 42
ctggacaagg agaaaaacat agggaaaaaa ccaacagaat ttgttggcat gttctacaca 60
cagaccatgg cttttcagaa gccaagctga ataaaaacag ttttaaaaga ggcaaccatt 120
tgtagaggag tccttgaagg attcttcatt gttttcttgg acaaaaagag accagtggat 180
ccaagtgctt caaatacttc tctcttattt tcttaactct attgctctgc aatatttact 240
ttaccctgtt aatgaacagg acaaaatggt taaaaaagag ataagcgtgc gtcaacaaat 300
tcaggctctt ctgtacaaga attttcttaa aaaatggaga ataaaaagag agtttctgga 360
ggaatggaca ataacattgt ttctagggct atatttgtgc atcttttcgg aacacttcag 420
agctacccgt tttcctgaac aacctcctaa agtcctggga agcgtggatc agtttaatga 480
ctctggcctg gtagtggcat atacaccagt cagtaacata acacaaagga taatgaataa 540
gatggccttg gcttccttta tgaaaggaag aacagtcatt gggacaccag atgaagagac 600
catggatata gaacttccaa aaaaatacca tgaaatggtg ggagttatat ttagtgatac 660
tttctcatat cgcctgaagt ttaattgggg atatagaatc ccagttataa aggagcactc 720
tgaatacaca ggtcactgtt gggccatgca tggtgaaatt ttttgttact tggcaaagta 780
ctggctaaaa gggtttgtag cttttcaagc tgcaattaat gctgcaatta tagaagtaac 840
aacaaatcat tctgtaatgg aggagttgac atcagttatt ggaataaata tgaagatacc 900
acctttcatt tctaagggag aaattatgaa tgaatggttt catttactt gcttagtttc 960
tttctcttct tttatatact ttgcatcatt aaatgttgca agggaaagag gaaaatttaa 1020
gaaactgatg acagtgatgg gtctccgaga gtcagcattc tggctctcct ggggattgac 1080
atacatttgc ttcatcttca ttatgtccat ttttatggct ctggtcataa catcaatccc 1140
aattgtattt catactggct tcatggtgat attcacactc tatagcttat atggcctttc 1200
tttggtggca ttggctttcc tcatgagtgt tttaataagg aaacctatgc tcgctggttt 1260
ggctggattt ctcttcactg tattttgggg atgtctggga ttcactgtgt tatatagaca 1320
acttccttta tctttgggat gggtattaag tcttcttagc ccttttgcct tcactgctgg 1380
aatggcccag attacacacc tggataatta cttaagtggt gttatttttc ctgatccctc 1440
tggggattca tacaaaatga tagccacttt tttcattttg gcatttgata ctcttttcta 1500
tttgatattc acattatatt ttgagcgagt tttacctggt aaggatggcc atggggattc 1560
tccattattt ttccttaagt cctcattttg gtccaaacat caaaatactc atcatgaaat 1620
ctttgagaat gaaataaatc ctgagcattc ctctgatgat tcttttgaac cggtgtctcc 1680
agaattccat ggaaaagaag ccataagaat cagaaatgtt ataaaagaat ataatggaaa 1740
gactggaaaa gtagaagcat tgcaaggcat attttttgac atatatgaag gacagatcac 1800
tgcaatactt gggcataatg gagctggtaa atcaacactg ctaaacattc ttagtggatt 1860
gtctgtttct acagaaggat cagccactat ttataatact caactctctg aaataactga 1920
catggaagaa attagaaaga atattggatt ttgtccacag ttcaattttc aatttgactt 1980
cctcactgtg agagaaaacc tcagggtatt tgctaaaata aaagggattc agccaaagga 2040
agtggaacaa gaggtattgc tgctagatga accaactgct ggattggatc ccttttcaag 2100
acaccgagtg tggagcctcc tgaaggagca taaagtagac cgacttatcc tcttcagtac 2160
ccaattcatg gatgaggctg acatcttggc tgataggaaa gtatttctgt ctaatgggaa 2220
gttgaaatgt gcaggatcat ctttgtttct gaagcgaaag tggggtattg gatatcattt 2280
aagtttacac aggaatgaaa tgtgtgacac agaaaaaatc acatcccttta ttaagcagca 2340

cattcctgat gccaagttaa caacagaaag tgaagaaaaa cttgtatata gtttgccttt 2400
ggaaaaaacg aacaaatttc cagatcttta cagtgacctt gataagtgtt ctgaccaggg 2460
cataaggaat tatgctgttt cagtgacatc tctgaatgaa gtattcttga acctagaagg 2520
aaaatcagca attgatgaac cagattttga cattgggaaa caagagaaaa tacatgtgac 2580
aagaaatact ggagatgagt ctgaaatgga acaggttctt tgttctcttc ctgaaacaag 2640
aaaggctgtc agtagtgcag ctctctggag acgacaaatc tatgcagtgg caacacttcg 2700
cttcttaaag ttaaggcgtg aaaggagagc tcttttgtgt ttgttactag tacttggaat 2760
tgcttttatc cccatcattc tagagaagat aatgtataaa gtaactcgtg aaactcattg 2820
ttgggagttt tcacccagta tgtatttcct ttctctggaa caaatcccga agacgcctct 2880
taccagcctg ttaatcgtta ataatacagg atcaaatatt gaagacctcg tgcattcact 2940
gaagtgtcag gatatagttt tggaaataga tgactttaga aacagaaatg gctcagatga 3000
tccctcctac aatggagcca tcatagtgtc tggtgaccag aaggattaca gattttctgt 3060
tgcgtgtaat accaagaaat tgaattgttt tcctgttctt atgggaattg ttagcaatgc 3120
ccttatggga atttttaact tcacggagct tattcaaatg gagagcactt catttttttt 3180
ttacataacc acaaaatctt ttcaaactaa gatcccttcc tctatccctt ctattctttg 3240
tcagaaaaat gttcaatccc agttatggat ttcaggcctc tggccttcag catactggtg 3300
tggacaggct ctggtggaca ttccattata cttcttgatt ctcttttcaa tacatttaat 3360
ttactacttc atatttctgg gattccagct ttcatgggaa ctcatgtttg ttttggtggt 3420
atgcataatt ggttgtgcag tttctcttat attcctcaca tatgtgcttt cattcatctt 3480
tcgcaagtgg agaaaaaata atggcttttg gtcttttggc ttttttattg taagtatata 3540
tacagacttt agctttcatt acaatgtttc taggtgtgat ttctatttta tctttatttt 3600
tgtatgttta tttattgctc atcattttc tttctgttct ccataccttc agagtgttat 3660
tttccttttt gtcataaggt gtctggaaat gaagtatgga aatgaaataa tgaataaaga 3720
cccagttttc agaatctctc cacggagtag agaaactcat cccaatccgg aagagcccga 3780
agaagaagat gaagatgttc aagctgaaag agtccaagca gcaaatgcac tcactgctcc 3840
aaacttggag gaggaaccag tcataactgc aagctgttta cacaaggaat attatgagac 3900
aaagaaaagt tgcttttcaa caagaaagaa gaaaatagcc atcagaaatg tttccttttg 3960
tgttaaaaaa ggtgaagttt tgggattact aggacacaat ggagctggta aaagtacttc 4020
cattaaaatg ataactgggt gcacaaagcc aactgcagga gtggtggtgt tacaaggcag 4080
cagagcatca gtaaggcaac agcatgacaa cagcctcaag ttcttggggt actgccctca 4140
ggagaactca ctgtggccca agcttacaat gaaagagcac ttggagttgt atgcagctgt 4200
gaaaggactg ggcaaagaag atgctgctct cagtatttca cgattggtgg aagctcttaa 4260
gctccaggaa caacttaagg ctcctgtgaa aactctatca gagggaataa agagaaagct 4320
gtgctttgtg ctgagcatcc tggggaaccc atcagtggtg cttctagatg agccgttcac 4380
cgggatggac cccgaggggc agcagcaaat gtggcagata cttcaggcta ccgttaaaaa 4440
caaggagagg ggcaccctct tgaccaccca ttacatgtca gaggctgagg ctgtgtgtga 4500
ccgtatggcc atgatggtgt caggaacgct aaggtgtatt ggttccattc aacatctgaa 4560
aaacaagttt ggtagagatt atttactaga aataaaaatg aaagaaccta cccaggtgga 4620
agctctccac acagagattt tgaagctttt cccacaggct gcttggcagg aaagatattc 4680
ctctttaatg gcgtataagt tacctgtgga ggatgtccac cctctatctc gggccttttt 4740
caagttagag gcgatgaaac agaccttcaa cctggaggaa tacagcctct ctcaggctac 4800
cttggagcag gtattcttag aactctgtaa agagcaggag ctgggaaatg ttgatgataa 4860
aattgataca acagttgaat ggaaacttct cccacaggaa gacccttaaa atgaagaacc 4920
tcctaacatt caattttagg tcctactaca ttgttagttt ccataattct acaagaatgt 4980
ttcctttac ttcagttaac aaaagaaaac atttaataaa cattcaataa tgattacagt 5040
tttcattttt aaaaatttag gatgaaggaa acaaggaaat atagggaaaa gtagtagaca 5100
aaattaacaa aatcagacat gttattcatc cccaacatgg gtctattttg tgcttaaaaa 5160
taatttaaaa atcatacaat attaggttgg ttatcggtta ttatcaataa agctaacact 5220
gagaacattt tacaaataaa aatatgaggt ttttagcctg aacttcaaat gtatcagcta 5280
tttttaaaca ttatttactc ggattctaat ttaatgtgac attgactata agaaggtctg 5340
ataaactgat gaaatggcac agcataacat ttaattataa tgacattctg attataaaat 5400
aaattgcatg tgaattttag tacatattga agttatatgg aagaagatag ccataatctg 5460
taagaaagta ccgcagttta atattttctt tagccaactt atattcatgt atttttatgg 5520
atctttttca aggtagtatc agtaggctag tcatttcgtt ctttcactca cgtcagaact 5580
tccatgtttt tgcctggttt atgggtagtt aggttggtac ca 5622

<210> 43
<211> 2600
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature

<223> Incyte ID No: 6952742CB1

<400> 43
cccgtccagt ggaccccctc acgatggatt ctggatccag cagggcccag cgtccatcca 60
taccgggcag gggctgggg cccgcgctgc caggagaagg cccagcacca atccccggac 120
ctgggtgggc gaggggtccg ccccaagggg cccgttgctg ccggggacct tgtcgtttgg 180
ccctggatcc gggggctcct gtgaccatgc cctcttctcg gccgcaggtc ggccacggga 240
cctgacgcaa caggatggac gagtcccctg agcctctgca gcagggcaga gggccggtgc 300
cggtccgacg ccagcgccca gcaccccggg gtctgcgtga gatgctgaag gccaggctgt 360
ggtgcagctg ctcgtgcagt gtgctgtgcg tccgggcgct ggtgcaggac ctgctccccg 420
ccacgcgctg gctgcgtcag taccgcccgc gggagtacct ggcaggcgac gtcatgtctg 480
ggctggtcat cggcatcatc ctggccatcg cctactcatt gctggccggg ctgcagccca 540
tctacagcct ctatacgtcc ttcttcgcca acctcatcta cttcctcatg ggcacctcac 600
ggcatgtctc cgtgggcatc ttcagcctgc tttgcctcat ggtggggcag gtggtggacc 660
gggagctcca gctggccggc tttgacccct cccaggacgg cctgcagccc ggagccaaca 720
gcagcaccct caacggctcg gctgccatgc tggactgcgg gcgtgactgc tacgccatcc 780
gtgtcgccac cgccctcacg ctgatgaccg ggctttacca ggtcctcatg ggcgtcctcc 840
ggctgggctt cgtgtccgcc tacctctcac agccactgct cgatggcttt gccatggggg 900
cctccgtgac catcctgacc tcgcagctca aacacctgct gggcgtgcgg atcccgcggc 960
accaggggcc cggcatggtg gtcctcacat ggctgagcct gctgcgcggc gccgggcagg 1020
ccaacgtgtg cgacgtggtc accagcacgg tgtgcctggc ggtgctgcta gccgcgaagg 1080
agctctcaga ccgctaccga caccgcctga gggtgccgct gcccacggag ctgctggtca 1140
tcgtggtggc cacactcgtg tcgcacttcg ggcagctcca caagcgcttt ggctcgagcg 1200
tggctggcga catccccacg ggtttcatgc cccctcaggt cccagagccc aggctgatgc 1260
agcgtgtggc tttggatgcc gtggccctgg ccctcgtggc tgccgccttc tccatctcgc 1320
tggcggagat gttcgcccgc agtcacggct actctgtgcg tgccaaccag gagctgctgg 1380
ctgtgggctg ctgcaacgtg ctacccgcct tcctccactg cttcgccacc agcgccgccc 1440
tggccaagag cctggtgaag acagccactg gctgccggac acagctgtcc agcgtggtca 1500
gcgccaccgt ggtgctgctg gtgctgctgg cgctggcacc gctgttccac gacctacagc 1560
gaagcgtgct ggcctgcgtc atcgtggtca gcctgcgggg ggccctgcgc aaggtgtggg 1620
acctcccgcg gttgtggcgg atgagcccgg ctgacgcgct ggtctgggca ggcaccgtgg 1680
ccacctgtat gctggtcagc acagaggccg ggctgctggc tggcgtcatc ctctcgctgc 1740
tcagcctggc cggccgcacc caaagccacg gcaccgccct gctggcccgc atcggggaca 1800
cggccttcta cgaggatgcc acagagttcg agggcctcgt ccctgagccc ggcgtgcggg 1860
tgttccgctt tgggggcccg ctgtactatg ccaacaagga cttcttcctg cagtcactct 1920
acagcctcac ggggctggac gcagggtgca tggctgccag gaggaaggag gggggctcag 1980
agacgggggt cggtgaggga ggccctgccc agggcgagga cctgggcccg gttagcacca 2040
gggctgcgct ggtgcccgca gcggccggct tccacacagt ggtcatcgac tgcgccccgc 2100
tgctgttcct agacgcagcc ggtgtgagca cgctgcagga cctgcgccga gactacgggg 2160
ccctgggcat cagcctgctg ctagcctgct gcagcccgcc tgtgagagac attctgagca 2220
gaggaggctt cctcggggag ggccccgggg acacggctga ggaggagcag ctgttcctca 2280
gtgtgcacga tgccgtgcag acagcacgag cccgccacag ggagctggag gccaccgatg 2340
cccatctgta gcagggccag gcctgcccag cagcctctgc tccctcctgg ggacccacag 2400
cagacgtctg caagccactg ctgagaccct tcccagggag gagccaccca agagctgcac 2460
tcttgtgcca cagctgccct ggggaaaccg gggaacccca actgggaaag gaggccctct 2520
gatcacacgc aggacccaaa cactcagaaa tcaagaacct ctgcctccga gacaggctgg 2580
cccacagtgc tggctgggcc 2600

<210> 44
<211> 2917
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7478795CB1

<400> 44
gcagcgggaa gagcggagcg aggaccgcgt ccggcgcagt cttcaatgag cagcgcggaa 60
actgcacccc agacccgagc ctgctgcgcg ccccctccca gagctcacct ggtgccaggt 120
aacaggcctg gcctcgccct gtggatgatg atggccttgc ccccgtgagc tacaacctgg 180
ccttcagcac ccgcccacct ccaaccagca ggatgcggct gtggaaggcg gtggtggtga 240
ctttggcctt catgagtgtg gacatctgcg tgaccacggc catctatgtc ttcagccacc 300

tggaccgcag cctcctggag gacatccgcc acttcaacat ctttgactcg gtgctggatc 360
tctgggcagc ctgcctgtac cgcagctgcc tgctgctggg agccaccatt ggtgtggcca 420
agaacagtgc gctggggccc cggcggctgc gggcctcgtg gctggtcatc accctcgtgt 480
gcctcttcgt gggcatctat gccatggtga agctgctgct cttctcagag gtgcgcaggc 540
ccatccggga cccctggttt tgggccctgt tcgtgtggac gtacatttca ctcggcgcat 600
ccttcctgct ctggtggctg ctgtccaccg tgcggccagg cacccaggcc ctggagccag 660
gggcggccac cgaggctgag ggcttccctg ggagcggccg gccaccgccc gagcaggcgt 720
ctggggccac gctgcagaag ctgctctcct acaccaagcc cgacgtggcc ttcctcgtgg 780
ccgcctcctt cttcctcatc gtggcagctc tgggagagac cttcctgccc tactacacgg 840
gccgcgccat tgatggcatc gtcatccaga aaagcatgga tcagttcagc acggctgtcg 900
tcatcgtgtg cctgctggcc attggcagct catttgccgc aggtattcgg ggcggcattt 960
ttaccctcat atttgccaga ctgaacattc gccttcgaaa ctgtctcttc cgctcactgg 1020
tgtcccagga gacaagcttc tttgatgaga accgcacagg ggacctcatc tcccgcctga 1080
cctcggacac caccatggtc agcgacctgg tctcccagaa catcaatgtc ttcctgcgga 1140
acacagtcaa ggtcacgggc gtggtggtct tcatgttcag cctctcatgg cagctctcct 1200
tggtcacctt catgggcttc cccatcatca tgatggtgtc caacatctac ggcaagtact 1260
acaagaggct ctccaaagag gtccagaatg ccctggccag agcgagcaac acggcggagg 1320
agaccatcag tgccatgaag actgtccgga gcttcgccaa tgaggaggag gaggcagagg 1380
tgtacctgcg gaagctgcag caggtgtaca agctgaacag gaaggaggca gctgcctaca 1440
tgtactacgt ctggggcagc gggctcacac tgctggtggt ccaggtcagc atcctctact 1500
acggggggcca ccttgtcatc tcaggccaga tgaccagcgg caacctcatc gccttcatca 1560
tctacgagtt tgtcctggga gattgtatgg agtccgtggg ctccgtctac agtggcctga 1620
tgcagggagt gggggctgct gagaaggtgt tcgagttcat cgaccggcag ccgaccatgg 1680
tgcacgatgg cagcttggcc cccgaccacc tggagggccg ggtggacttt gagaatgtga 1740
ccttcaccta ccgcactcgg ccccacaccc aggtcctgca gaatgtctcc ttcagcctgt 1800
cccccggcaa ggtgacggcc ctggtggggc cctcgggcag tgggaagagc tcctgtgtca 1860
acatcctgga gaacttctac cccctggagg ggggccgggt gctgctggac ggcaagccca 1920
tcagcgccta cgaccacaag tacttgcacc gtgtgatctc cctggtgagc caggagcccg 1980
tgctgttcgc ccgctccatc acggataaca tctcctacgg cctgcccact gtgcctttcg 2040
agatggtggt ggaggccgca cagaaggcca atgcccacgg cttcatcatg gaactccagg 2100
acggctacag cacagagaca ggggagaagg gcgcccagct gtcaggtggc cagaagcagc 2160
gggtggccat ggcccgggct ctggtgcgga accccccagt cctcatcctg gatgaagcca 2220
ccagcgcttt ggatgccgag agcgagtatc tgatccagca ggccatccat ggcaacctgc 2280
agaagcacac ggtactcatc atcgcgcacc ggctgagcac cgtggagcac gcgcacctca 2340
ttgtggtgct ggacaagggc cgcgtagtgc agcagggcac ccaccagcag ctgctggccc 2400
agggcggcct ctacgccaag ctggtgcagc ggcagatgct ggggcttcag cccgccgcag 2460
acttcacagc tggccacaac gagcctgtag ccaacggcag tcacaaggcc tgatgggggg 2520
cccctgcttc tcccggtggg gcagaggacc cggtgcctgc ctggcagatg tgcccacgga 2580
ggcccccagc tgccctccga gcccaggcct gcagcactga aagacgacct gccatgtccc 2640
atggatcacc gcttcctgca tcttgcccct ggtccctgcc ccattcccag ggcactcctt 2700
accccctgctg ccctgagcca acgccttcac ggacctccct agcctcctaa gcaaaggtag 2760
agctgccttt ttaaacctag gtcttaccag ggtttttact gtttggtttg aggcacccca 2820
gtcaactcct agatttcaaa aacctttttc taattgggag taatggcggg cactttcacc 2880
aagatgttct agaaacttct gagccaggag tgaatgg 2917

<210> 45
<211> 1474
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 656293CB1

<400> 45
atggggctcc ggagccacca cctcagcctg ggccttctgc ttctgtttct actccctgca 60
gagtgcctgg gagctgaggg ccggctggct ctcaagctgt tccgtgacct ctttgccaac 120
tacacaagtg ccctgagacc tgtggcagac acagaccaga ctctgaatgt gaccctggag 180
gtgacactgt cccagatcat cgacatggat gaacggaacc aggtgctgac cctgtatctg 240
tggatacggc aggagtggac agatgcctac ctacgatggg accccaatgc ctatggtggc 300
ctggatgcca tccgcatccc cagcagtctt gtgtggcggc cagacatcgt actctataac 360
aaagccgacg cgcagcctcc aggttccgcc agcaccaacg tggtcctgcg ccacgatggc 420
gccgtgcgct gggacgcgcc ggccatcacg cgcagctcgt gccgcgtgga tgtagcagcc 480

ttcccgttcg acgcccagca ctgcggcctg acgttcggct cctggactca cggcgggcac 540
caactggatg tgcggccgcg cggcgctgca gccagcctgg cggacttcgt ggagaacgtg 600
gagtggcgcg tgctgggcat gccggcgcgg cggcgcgtgc tcacctacgg ctgctgctcc 660
gagccctacc ccgacgtcac cttcacgctg ctgctgcgcc gccgcgccgc cgcctacgtg 720
tgcaacctgc tgctgccctg cgtgctcatc tcgctgcttg cgccgctcgc cttccacctg 780
cctgccgact caggcgagaa ggtgtcgctg ggcgtcaccg tgctgctggc gctcaccgtc 840
ttccagttgc tgctggccga gagcatgcca ccggccgaga gcgtgccgct catcgggaag 900
tactacatgg ccactatgac catggtcaca ttctcaacag cactcaccat ccttatcatg 960
aacctgcatt actgtggtcc cagtgtccgc ccagtgccag cctgggctag ggccctcctg 1020
ctgggacacc tggcacgggg cctgtgcgtg cgggaaagag gggagccctg tgggcagtcc 1080
aggccacctg agttatctcc tagcccccag tcgcctgaag gaggggctgg ccccccagcg 1140
ggcccttgcc acgagccacg atgtctgtgc cgccaggaag ccctactgca ccacgtagcc 1200
accattgcca ataccttccg cagccaccga gctgcccagc gctgccatga ggactggaag 1260
cgcctggccc gtgtgatgga ccgcttcttc ctggccatct tcttctccat ggccctggtc 1320
atgagcctcc tggtgctggt gcaggccctg tgagggctgg gactaagtca cagggatctg 1380
ctgcagccac agctcctcca gaaagggaca gccacggcca agtggttgct ggtctttggg 1440
ccagccagtc tctccccact gctcctaaga tcct 1474

<210> 46
<211> 1742
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7473957CB1

<400> 46
tgggctcggc tccctgcctc cgcgtcgcag cccccgccgt agccgcctcc gagcccgccg 60
ccacatcctc tgagcagaag atggctgtgc cacccacgta tgccgatctt ggcaaatctg 120
ccagggatgt cttcaccaag ggctatggaa tttacaagct caggctcagc caacactgag 180
accaccaaag tgacgggcag tctggaaacc aagtacagat ggactgagta cggcctgacg 240
tttacagaga aatggaatac cgacaataca ctaggcaccg agattactgt ggaagatcag 300
cttgcacgtg gactgaagct gaccttcgat tcatccttct cacctaacac tgggaaaaaa 360
aatgctaaaa tcaagacagg gtacaagcgg gagcacatta acctgggctg cgacatggat 420
ttcgacattg ctgggccttc catccggggt gctctggtgc taggttacga gggctggctg 480
gccggctacc agatgaattt tgagactgca aaatcccgag tgacccagag caactttgca 540
gttggctaca agactgatga attccagctt cacactaatg tgaatgacgg gacagagttt 600
ggcggctcca tttaccagaa agtgaacaag aagttggaga ccgctgtcaa tcttgcctgg 660
acagcaggaa acagtaacac gcgcttcgga atagcagcca agtatcagat tgaccctgac 720
gcctgcttct cggctaaagt gaacaactcc agcctgatag gtttaggata cactcagact 780
ctaaagccag gtattaaact gacactgtca gctcttctgg atggcaagaa cgtcaatgct 840
ggtggccaca agcttggtct aggactggaa tttcaagcat aaatgaatac tgtacaattg 900
tttaatttta aactattttg cagcatagct accttcagaa tttagtgtat cttttaatgt 960
tgtatgtctg ggatgcaagt attgctaaat atgttagccc tccaggttaa agttgattca 1020
gctttaagat gttacccttc cagaggtaca gaagaaacct atttccaaaa aaggtccttt 1080
cagtggtaga ctcggggaga acttggtggc ccctttgaga tgccaggttt cttttttatc 1140
tagaaatggc tgcaagtgga agcggataat atgtaggcac tttgtaaatt catattgagt 1200
aaatgaatga aattgtgatt tcctgagaat cgaaccttgg ttccctaacc ctaattgatg 1260
agaggctcgc tgcttgatgg tgtgtacaaa ctcacctgaa tgggactttt ttagacagat 1320
cttcatgacc tgttcccacc ccagttcatc atcatctctt ttacaccaaa aggtctgcag 1380
ggtgtggtaa ctgtttcttt tgtgccattt tggggtggag aaggtggatg tgatgaagcc 1440
aataattcag gacttattcc ttcttgtgtt gtgtttttttt ttggcccttg caccagagta 1500
tgaaatagct tccaggagct ccagctataa gcttggaagt gtctgtgtga ttgtaatcac 1560
atggtgacaa cactcagaat ctaaattgga cttctgttgt attctcacca ctcaatttgt 1620
tttttagcag ttaatgggta catttttagag tctccatttt gttggaatta gatcctccct 1680
tcaagggctg gattacacac ttaaaaactg attaatattg acttaaaaaa aagggccgca 1740
ta 1742

<210> 47
<211> 2312
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7474111CB1

<400> 47
ccaagagaaa ggctgttttt gccggtgaca aggcgttcct ccccagactt tcatcccact 60
ctaagggcag catctcgaga gggtccaggc gcgctacacg ttaggcgcct tctcaggact 120
cgcgccccag aacgtggggg ccggggccgg gtcggggacc gcttccctcc gcgggctggc 180
aggcggcacc gaggctgggg gatggggcgt gtagccgccc tctgagcacg cggtacgtgg 240
gtcgccccgc gcggcggtag gaagagcaaa gtcggcagga aaagccgtgg ctgggatgcc 300
ttccctgaga aatccttagg ggcgatgtca gaacccgagc tgggctccgg acagtttctg 360
gaaaaagctc tccaaacgcc gtctgtcccc gcacccgagt ccacactggg ctttgaacca 420
gggctgttaa aaggggccct ggggactgcc caattcatcc cgatggccca gggcaggacg 480
cgggagcagg catcccggcg ctgggctccc cgctcccccg ccctgcggac ccctccccgc 540
cactacgggc cggagcggag ggggaggacg gcgtcacgag gcggggagcc cgaggtccag 600
ggcggggcgc ccgggaatcc cagcccgagc aagccgggga gccctcaggg ggtcggcccc 660
gcggcttggg agagggcacc gcggcctcgg tgtgcgcagc cctcgggcgc gagggtcggc 720
gagcggacac agccgcgttc ccagccggtg gggctcagcc gtggcgccgg cgaggactcc 780
ccggccaccc gcagcggggc ggcctcggtg gtgctgaacg tgggcggcgc ccggtattcg 840
ctgtcccggg agctgctgaa ggacttcccg ctgcgccgcg tgagccggct gcacggctgc 900
cgctccgagc gcgacgtgct cgaggtgtgc gacgactacg accgcgagcg caacgagtac 960
ttcttcgacc ggcactcgga ggccttcggc ttcatcctgc tctacgcggc tccctccagg 1020
cgctggctgg agcgcatgcg gcggaccttc gaggagccca cgtcgtcgct ggccgcgcag 1080
atcctggcta gcgtgtcggt ggtgttcgtg atcgtgtcca tggtggtgct gtgcgccagc 1140
acgttgcccg actggcgcaa cgcagccgcc gacaaccgca gcctggatga ccggagcagg 1200
ataattgaag ctatctgcat aggttggttc actgccgagt gcatcgtgag gttcattgtc 1260
tccaaaaaca agtgtgagtt tgtcaagaga cccctgaaca tcattgattt actggcaatc 1320
acgccgtatt acatctctgt gttgatgaca gtgtttacag gcgagaactc tcaactccag 1380
agggctggag tcaccttgag ggtacttaga atgatgagga ttttttgggt gattaagctt 1440
gcccgtcact tcattggtct tcagacactc ggtttgactc tcaaacgttg ctaccgagag 1500
atggttatgt tacttgtctt catttgtgtt gccatggcaa tctttagtgc actttctcag 1560
cttcttgaac atgggctgga cctggaaaca tccaacaagg actttaccag cattcctgct 1620
gcctgctggt gggtgattat ctctatgact acagttggct atggagatat gtatcctatc 1680
acagtgcctg gaagaattct tggaggagtt tgtgttgtca gtggaattgt tctattggca 1740
ttacctatca cttttatcta ccatagcttt gtgcagtgtt atcatgagct caagtttaga 1800
tctgctaggt atagtaggag cctctccact gaattcctga attaatgcat tgcaaatcaa 1860
ttcttgcata cacttcatag aaagactttg atgctgcttc atatttatgt gtttcttgct 1920
gggtgagcac tgcagtggca ttgtcatcat cttggtaggg taaaaattat ccttcccagc 1980
cgaagggata aaacagttta cttgttatgg agtaaataga attgagactg caaaggaaga 2040
ataatgactc ctagagtaaa ctttaggacc cggttttatt tagacttgtt ttcccgtttc 2100
cttgaatgat tacacatttt taaaaaatac attatttgaa cattttaaaa cagaaaggta 2160
ctattttcca aatgtttttc catcttatga attcagaaga agcttggaac ttatagtgtt 2220
ttttgtttga gagtaacatt ttcatttcta aatgttttat aatttctcat atcaatgtca 2280
gaagtatcct ggaaacatat gtcacatgcg ag 2312

<210> 48
<211> 2320
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7480826CB1

<400> 48
cccgccgctc gcagggctgc tccacagccg cgcgacgccg ccgccttaga acgcctttcc 60
agtactgcta gcagcagccc gaccacgcgt taccgcacgc tcgcgccttt cccttgacac 120
ggcggacgcc ggaggattgg ggcggcaatt tgtcttttcc tttttatta aaattatttt 180
tcctgcctgt tgttggattt ggggaaattt tttgtttgtt ttttatgatt tgtatttgac 240
tgagagaaac ccactgaaga cgtctgcgtg agaatagaga ccaccgaggc cgactcgcgg 300
gccgctgcac ccaccgccaa ggacaaaagg agcccagcgc tactagctgc acccgattcc 360
tcccagtgct tagcatgaag aaggccgaaa tgggacgatt cagtatttcc ccggatgaag 420
acagcagcag ctacagttcc aacagcgact tcaactactc ctaccccacc aagcaagctg 480

ctctgaaaag ccattatgca gatgtagatc ctgaaaacca gaacttttta cttgaatcga 540
atttggggaa gaagaagtat gaaacagaat ttcatccagg tactacttcc tttggaatgt 600
cagtatttaa tctgagcaat gcgattgtgg gcagtggaat ccttgggctt tcttatgcca 660
tggctaatac tggaattgct ctttttataa ttctcttgac atttgtgtca atattttccc 720
tgtattctgt tcatctcctt ttgaagactg ccaatgaagg agggtcttta ttatatgaac 780
aattgggata taaggcattt ggattagttg gaaagcttgc agcatctgga tcaattacaa 840
tgcagaacat tggagctatg tcaagctacc tcttcatagt gaaatatgag ttgcctttgg 900
tgatccaggc attaacgaac attgaagata aaactggatt gtggtatctg aacgggaact 960
atttggttct gttggtgtca ttggtggtca ttcttccttt gtcgctgttt agaaatttag 1020
gatatttggg atataccagt ggcctttcct tgttgtgtat ggtgttcttt ctgattgtgg 1080
tcatttgcaa gaaatttcag gttccgtgtc ctgtggaagc tgctttgata attaacgaaa 1140
caataaacac caccttaaca cagccaacag ctcttgtacc tgctttgtca cataacgtga 1200
ctgaaaatga ctcttgcaga cctcactatt ttattttcaa ctcacagact gtctatgctg 1260
tgccaattct gatcttttca tttgtctgtc atcctgctgt tcttcccatc tatgaagaac 1320
tgaaagaccg cagccgtaga agaatgatga atgtgtccaa gatttcattt tttgctatgt 1380
ttctcatgta tctgcttgcc gccctctttg gatacctaac attttacgaa catgttgagt 1440
cagaattgct tcatacctac tcttctatct tgggaactga tattcttctt ctcattgtcc 1500
gtctggctgt gttaatggct gtgaccctga cagtaccagt agttattttc ccaatccgga 1560
gttctgtaac tcacttgttg tgtgcatcaa aagatttcag ttggtggcgt catagtctca 1620
ttacagtgtc tatcttggca tttaccaatt tacttgtcat ctttgtccca actattaggg 1680
atatctttgg ttttattggt gcatctgcag cttctatgtt gatttttatt cttccttctg 1740
ccttctatat caagttggtg aagaaagaac ctatgaaatc tgtacaaaag attggggctt 1800
tgttcttcct gttaagtggt gtactggtga tgaccggaag catggccttg attgttttgg 1860
attgggtaca caatgcacct ggaggtggcc attaattggc accactcaaa ctcaaactca 1920
gtccatctga tgccagtgtt gagtaaactc aactactatg aaatttcacc taatgttttc 1980
agtttcactt ccttttgaag tgcagattcc tcgctggttc ttctgagtgc agaataagtg 2040
aactttttg ttttgttttg tttttttaag aaacttatct gtatgttaga aatggatatg 2100
aacaacaaaa ccacgagtct cgggttaagg gaagtgacaa ttttattcca ttccagagaa 2160
tggacaaact cttaactttt atcaagccac atgcttggct gtgtcattgt ttaacttgga 2220
tattttatga ttttacttga atgtgcctaa tggaaccatt tgatgtgaga aacaattctt 2280
tttaatttac agcaaaatat tgaataacca ttgacaaaaa 2320

<210> 49
<211> 1781
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 6025572CB1

<400> 49
gtcactttct cgccagtacg atgctgcagc ggttttccgg ttttccgctt cccttcatcg 60
tagctcccgt actcattttt agccactgct gccggttttt atatccttct ccatcatgca 120
tcgtgagcct gcgaaaaaga aggcagaaaa gcggctgttt gacgcctcat ccttcgggaa 180
ggaccttctg gccggcggag tcgcggcagc tgtgtccaag acagcggtgg cgcccatcga 240
gcgggtgaag ctgctgctgc aggtgcaggc gtcgtcgaag cagatcagcc ccgaggcgcg 300
gtacaaaggc atggtggact gcctggtgcg gattcctcgc gagcagggtt tcttcagttt 360
ttggcgtggc aatttggcaa atgttattcg gtattttcca acacaagctc taaactttgc 420
ttttaaggac aaatacaagc agctattcat gtctggagtt aataaagaaa aacagttctg 480
gaggtggttt ttggcaaacc tggcttctgg tggagctgct ggggcaacat ccttatgtgt 540
agtatatcct ctagattttg cccgaacccg attaggtgtc gatattggaa aaggtcctga 600
ggagcgacaa ttcaagggtt taggtgactg tattatgaaa atagcaaaat cagatggaat 660
tgctggttta taccaagggt ttggtgtttc agtacagggc atcattgtgt accgagcctc 720
ttattttgga gcttatgaca cagttaaggg tttattacca aagccaaaga aaactccatt 780
tcttgtctcc tttttcattg ctcaagttgt gactacatgc tctggaatac tttcttatcc 840
ctttgacaca gttagaagac gtatgatgat gcagagtggt gaggctaaac ggcaatataa 900
aggaacctta gactgctttg tgaagatata ccaacatgaa ggaatcagtt cctttttttcg 960
tggcgccttc tccaatgttc ttcgcggtac agggggtgct ttggtgttgg tattatatga 1020
taaaattaaa gaattctttc atattgatat tggtggtagg taatcgggag agtaaattaa 1080
gaaatacatg gatttaactt gttaaacata caaattacat agctgccatt tgcatacatt 1140
ttgatagtgt tattgtctgt attttgttaa agtgctagtt ctgcaataaa gcatacattt 1200
tttcaagaat ttaaatacta aaaatcagat aaatgtggat tttcctccca cttagactca 1260

aacacatttt agtgtgatat ttcatttatt ataggtagta tattttaatt tgttagttta 1320
aaattctttt tatgattaaa aattaatcat ataatcctag attaatgctg aaatctagga 1380
aatgaaagta gcgtctttta aattgctatt catttaatat acctgtttc ccatcttttg 1440
aagtcatatg gtatgacata tttcttaaaa gcttatcaat agatgtcatc atatgtgtag 1500
gcagaaataa gctttgttct atatctcttc taagacagtt gttattactg tgtataatat 1560
ttacagtatc agcctttgat tatagatgtg atcatttaaa atttgataat gactttagtg 1620
acattataaa actgaaactg gaaaataaaa tggcttatct gctgatgttt atctttaaaa 1680
taaataaaat cttgctagtg tgaatatatc ttagaacaaa aggtatcctc ttgaaaatta 1740
gtttgtatat tttgttgaca ataaaggaag cttaactgtt a 1781

<210> 50
<211> 2433
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 5686561CB1

<400> 50
catccgctca caatgccaca tcaatgatac gagcacgtag cctcactgct tgcacagtgc 60
atggcagagt cggctgcgag caggcgaggt ggcctgaggg aggtcactag gctggctgag 120
ggctttttgc tgtggttctg agccggcctg cttccaggca ccgtgtccat gcgggtaagc 180
ggtctccctg ggtgcccact cttgcgcccg gagatcctga gtttggtcct gtctggccat 240
gaagctcagc ctgctgggag gccacaggga gatgcaggct gggcggcggg tggatggttc 300
cagccggttg ggtccggggc ctggagctca gcctgtgggg tggggaccca gtggtgccct 360
ggagctgccg cttctgctct cagcaggatg atgggcagga cagggagagg ctgacctact 420
tccagaacct gcctgagtct ctgacttccc tcctggtgct gctgaccacg gccaacaacc 480
ccgatgtgat gattcctgcg tattccaaga accgggccta tgccatcttc ttcatagtct 540
tcactgtgat aggaagcctg tttctgatga acctgctgac agccatcatc tacagtcagt 600
tccggggcta cctgatgaaa tctctccaga cctcgctgtt tcggaggcgg ctgggaaccc 660
gggctgcctt tgaagtccta tcctccatgg tgggggaggg aggagccttc cctcaggcag 720
ttggggtgaa gccccagaac ttgctgcagg tgcttcagaa ggtccagctg gacagctccc 780
acaaacaggc catgatggag aaggtgcgtt cctacgacag tgttctgctg tcagctgagg 840
agtttcagaa gctcttcaac gagcttgaca gaagtgtggt taaagagcac ccgccgaggc 900
ccgagtacca gtctccgttt ctgcagagcg cccagttcct cttcggccac tactactttg 960
actacctggg gaacctcatc gccctggcaa acctggtgtc catttgcgtg ttcctggtgc 1020
tggatgcaga tgtgctgcct gctgagcgtg atgacttcat cctggggatt ctcaactgcg 1080
tcttcattgt gtactacctg ttggagatgc tgctcaaggt ctttgccctg ggcctgcgag 1140
ggtacctgtc ctaccccagc aacgtgtttg acgggctcct caccgttgtc ctgctggttt 1200
tggagatctc aactctggct gtgtaccgat tgccacaccc aggctggagg ccggagatgg 1260
tgggcctgct gtcgctgtgg gacatgaccc gcatgctgaa catgctcatc gtgttccgct 1320
tcctgcgtat catccccagc atgaagccga tggccgtggt ggccagtacc gtcctgggcc 1380
tggtgcagaa catgcgtgct tttggcggga tcctggtggt ggtctactac gtatttgcca 1440
tcattgggat caacttgttt agaggcgtca ttgtggctct tcctggaaac agcagcctgg 1500
cccctgccaa tggctcggcg ccctgtggga gcttcgagca gctggagtac tgggccaaca 1560
acttcgatga ctttgcggct gccctggtca ctctgtggaa cttgatggtg gtgaacaact 1620
ggcaggtgtt tctggatgca tatcggcgct actcaggccc gtggtccaag atctattttg 1680
tgttgtggtg gctggtgtcg tctgtcatct gggtcaacct gtttctggcc ctgattctgg 1740
agaacttcct tcacaagtgg gacccccgca gccacctgca gccccttgct gggaccccag 1800
aggccaccta ccagatgact gtggagctcc tgttcaggga tattctggag gagcccgggg 1860
aggatgagct cacagagagg ctgagccagc acccgcacct gtggctgtgc aggtgacgtc 1920
cgggctgccg tcccagcagg ggcggcagga gagagaggct ggcctacaca ggtgcccatc 1980
atggaagagg cggccatgct gtggccagcc aggcaggaag agacctttcc tctgacggac 2040
cactaagctg gggacaggaa ccaagtcctt tgcgtgtggc ccaacaacca tctacagaac 2100
agctgctggt gcttcaggga ggcgccgtgc cctccgcttt cttttatagc tgcttcagtg 2160
agaattccct cgtcgactcc acagggacct ttcagacaaa aatgcaagaa gcagcggcct 2220
cccctgtccc ctgcagcttc cgtggtgcct ttgctgccgg cagcccttgg ggaccacagg 2280
cctgaccagg gcctgcacag gttaaccgtc agacttccgg ggcattcagg tggggatgct 2340
ggtggtttga catggagaga accttgactg tgttttatta tttcatggct tgtatgagtg 2400
tgactgggtg tgtttcttta gggttctgat tgc 2433

<210> 51

<211> 1772
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1553725CB1

<400> 51
cccacgcgtc cgaactggtg gcatttgtcc cgggaccagg tccacagttt tatgtgtgag 60
caagatggag gctgacctgt ctggctttaa catcgatgcc ccccgttggg accagcgcac 120
cttcctgggg agagtgaagc acttcctaaa catcacggac cccgcactg tctttgtatc 180
tgagcgggag ctggactggg ccaaggtgat ggtggagaag agcaggatgg gggttgtgcc 240
cccaggcacc caagtggagc agctgctgta tgccaagaag ctgtatgact cggccttcca 300
ccccgacact ggggagaaga tgaatgtcat cggcgcatg tctttccagc ttcctggcgg 360
catgatcatc acgggcttca tgctccagtt ctacaggacg atgccggcgg tgatcttctg 420
gcagtgggtg aaccagtcct tcaatgcctt agtcaactac accaacagga atgcggcttc 480
ccccacatca gtcaggcaga tggccctttc ctacttcaca gccacaacca ctgctgtggc 540
cacggctgtg ggcatgaaca tgttgacaaa gaaagcgccg cccttggtgg gccgctgggt 600
gcccttgcc gctgtggctg cggctaactg tgtcaatatc cccatgatgc gacagcagga 660
gctcataaag ggaatctgcg tgaaggacag gaatgaaaat gagattggtc attcccggag 720
agctgcggcc ataggcatca cccaagtagt tatttctcgg atcaccatgt cagctcctgg 780
gatgatcttg ctgccagtca tcatggaaag gcttgagaaa ttgcacttca tgcagaaagt 840
caaggtcctg cacgccccat tgcaggtcat gctgagcggg tgcttcctca tcttcatggt 900
gccagtggcg tgtgggcttt tcccacagaa atgtgaattg ccagtttcct atctggaacc 960
gaagctccaa gacactatca aggccaagta tggagaactt gagccttatg tctacttcaa 1020
taagggtctc taaatgcccc acttcagcaa ggaccagtct attcccatat tcaccagctc 1080
ctccttagct acgtgcacac ttgtgtcctc cttcccсttt gccaacaagg cctgaaggcc 1140
agggtagatt ggggggtggg acaatgaatg cctcatactt acaccctggt actggttgat 1200
tggacctcag gggaaaaaag tgaaaaaggg tagcaaaggc caatgtcttc tagctgcttc 1260
ctcaacccct gtcccctgga gaccagaagc tgaggccctc tcagggagga gacatccaag 1320
caaatcattt ggaaaagtta ggaaaccttt aggattctgg ttccagccag ggttgaggaa 1380
aagaccttgg atcaaaagga agcttctata cctctttctt cttcgcttcc tcctctccca 1440
agcaatggaa acttttaccc atgtaattct agctgaactc aggaaaaaga agggggaaag 1500
gactctgtcc ccttggggct catcaccctt ccacatcctc ctcctcgttg ccccctggtc 1560
aggcagcttc tttttttttt ttcaagatgg agtcttgctc tgtcgcccag gctggaatgc 1620
agtggcgcga tctcggctca ctgcaaactc tgcctcctgg attcaagcga ttctcctgcc 1680
tcagcctctc aagtagctgg gattacaggg cacctgccac cacgcctggc taatttttgt 1740
attttagtgg agacggggtt tcaccatgct gg 1772

<210> 52
<211> 1874
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1695770CB1

<400> 52
atcttttcca gctcccagct ggcaggctaa ggccctccgg agcccaaggc gagcccaagc 60
agaagccagt agggttatct gtgtcaggat catttccagg ggaatagttc tggcccctgg 120
caggtaaaga taggccagag gagaagaggc agaagaggag agaaagcagg ctcttttgcg 180
agcagcccag gttggagaaa ggctctgtac ttttggcgtt cctgcaggga tatcccctct 240
cacattggca gccaggctga gaaagggctt caagatcccc gcagaatgac aactcttgtt 300
cctgcaaccc tctccttcct tcttctctgg accctgccag ggcaggtcct cctcagggtg 360
gccttggcaa aagaggaagt caaatctgga accaaggggt cccagcccat gtccccctct 420
gatttcctag acaaacttat ggggcgaaca tctggatatg atgccaggat tcggcccaat 480
tttaaaggcc cacccgtgaa cgtgacctgc aacatcttca tcaacagttt cagctccgtc 540
accaagacca caatggacta ccgggtgaat gtcttcttgc ggcaacagtg gaatgaccca 600
cgcctgtcct accgagaata tcctgatgac tctctggacc tcgatccctc catgctggac 660
tctatctgga agccagacct cttctttgct aatgagaaag gggccaactt ccatgaggtg 720
accacggaca acaagttact gcgcatcttc aagaatggga atgtgctgta cagcatcagg 780

ctgaccctca ttttgtcctg cctgatggac ctcaagaact tccccatgga catccagacg 840
tgcacgatgc agcttgagag ctttggctac accatgaaag acctcgtgtt tgagtggctg 900
gaagatgctc ctgctgtcca agtggctgag ggctgactc tgccccagtt tatcttgcgg 960
gatgagaagg atctaggctg ttgtaccaag cactacaaca cagggaaatt cacctgcatc 1020
gaggtaaagt ttcacctgga acggcagatg ggctactatc tgattcagat gtacatcccc 1080
agcctactca tcgtcatcct gtcctgggtc tccttctgga tcaacatgga tgctgctcct 1140
gcccgtgtgg gcctgggcat caccaccgtg ctcaccatga ccacccagag ctctggctcc 1200
cgggcctctt tgcctaaggt gtcctacgtg aaggcaatcg acatctggat ggctgtgtgt 1260
ctgctctttg tgttcgctgc cttgctggag tatgctgcca taaattttgt ttctcgtcag 1320
cataaagaat tcatacgact tcgaagaagg cagaggcgcc aacgcttgga ggaagatatc 1380
atccaagaaa gtcgtttcta tttccgtggc tatggcttgg gccactgcct gcaggcaaga 1440
gatggaggtc caatggaagg ttctggcatt tatagtcccc aacctccagc ccctcttcta 1500
agggaaggag aaaccacgcg gaaactctac gtggactgag ccaagagaat tgacaccatc 1560
tcccgggctg tcttcccttt cactttcctc atcttcaata tcttctactg ggttgtctat 1620
aaagtgctac ggtcagaaga tatccaccag gctctgtgaa tagggtggga gctatagagt 1680
cctgctgctg gcctcctgct tcctcctggg tgggctttct ccctcagtta gactccatta 1740
ggggtttgga cagttccttc ctgatctccc actcagaact tcaactacca gtcccaaagc 1800
tatgtgggcc tatattgcat ggtgccaatg gtggctgtac ttataaagat ggttatctac 1860
ccttaaaaaa aaaa 1874

<210> 53
<211> 6211
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 4672222CB1

<400> 53
cggcggaggc gggcgcgggc gcgtccctgt ggccagtcac ccggaggagt tggtcgcaca 60
attatgaaag actcggcttc tgctgctagc gccggagctg agttagttct gagaaggttt 120
ccctgggcgt tccttgtccg gcggcctctg ctgccgcctc cggagacgct tcccgataga 180
tggctacagg ccgcggagga ggaggaggtg gagttgctgc ccttccggag tccgcccccgt 240
gaggagaatg tcccagaaat cctggataga aagcactttg accaagaggg aatgtgtata 300
tattatacca agttccaagg accctcacag atgccttcca ggatgtcaaa tttgtcagca 360
actcgtcagg tgtttttgtg gtcgcttggt caagcaacat gcttgtttta ctgcaagtct 420
tgccatgaaa tactcagatg tgaaattggg tgaccatttt aatcaggcaa tagaagaatg 480
gtctgtggaa aagcatacag aacagagccc aacggatgct tatggagtca taaattttca 540
agggggttct cattcctaca gagctaagta tgtgaggcta tcatatgaca ccaaacctga 600
agtcattctg caacttctgc ttaaagaatg gcaaatggag ttacccaaac ttgttatctc 660
tgtacatggg ggcatgcaga aatttgagct tcacccacga atcaagcagt tgcttggaaa 720
aggtcttatt aaagctgcag ttacaactgg agcctggatt ttaactggag gagtaaacac 780
aggtgtggca aaacatgttg gagatgccct caaagaacat gcttccagat catctcgaaa 840
gatttgcact atcggaatag ctccatgggg agtgattgaa aacagaaatg atcttgttgg 900
gagagatgtg gttgctcctt atcaaacctt attgaacccc ctgagcaaat tgaatgtttt 960
gaataatctg cattcccatt tcatattggt ggatgatggc actgttggaa agtatggggc 1020
ggaagtcaga ctgagaagag aacttgaaaa aactattaat cagcaaagaa ttcatgctag 1080
gattggccag ggtgtccctg tggtggcact tatatttgag ggtgggccaa atgttatcct 1140
cacagttctt gaataccttc aggaaagccc ccctgttcca gtagttgtgt gtgaaggaac 1200
aggcagagct gcagatctgc tagcgtatat tcataaacaa acagaagaag gagggaatct 1260
tcctgatgca gcagagcccg atattatttc cactatcaaa aaaacattta actttggcca 1320
gaatgaagca cttcatttat ttcaaacact gatggagtgc atgaaaagaa aggagcttat 1380
cactgttttc catattgggt cagatgaaca tcaagatata gatgtagcaa tacttactgc 1440
actgctaaaa ggtactaatg catctgcatt tgaccagctt atccttacat tggcatggga 1500
tagagttgac attgccaaaa atcatgtatt tgtttatgga cagcagtggc tggttggatc 1560
cttggaacaa gctatgcttg atgctcttgt aatggataga gttgcatttg taaaacttct 1620
tattgaaaat ggagtaagca tgcataaatt ccttaccatt ccgagactgg aagaacttta 1680
caacactaaa caaggtccaa ctaatccaat gctgtttcat cttgttcgag acgtcaaaca 1740
gggaaatctt cctccaggat ataagatcac tctgattgat ataggacttg ttattgaata 1800
tctcatggga ggaacctaca gatgcaccta tactaggaaa cgttttcgat taatatataa 1860
tagtcttggt ggaaataatc ggaggtctgg ccgaaatacc tccagcagca ctcctcagtt 1920
gcgaaagagt catgaatctt ttggcaatag ggcagataaa aaggaaaaaa tgaggcataa 1980

```
ccatttcatt aagacagcac agccctaccg accaaagatt gatacagtta tggaagaagg 2040
aaagaagaaa agaaccaaag atgaaattgt agacattgat gatccagaaa ccaagcgctt 2100
tccttatcca cttaatgaac ttttaatttg ggcttgcctt atgaagaggc aggtcatggc 2160
ccgtttttta tggcaacatg gtgaagaatc aatggctaaa gcattagttg cctgtaagat 2220
ctatcgttca atggcatatg aagcaaagca gagtgacctg gtagatgata cttcagaaga 2280
actaaaacag tattccaatg attttggtca gttggccgtt gaattattag aacagtcctt 2340
cagacaagat gaaaccatgg ctatgaaatt gctcacttat gaactgaaga actggagtaa 2400
ttcaacctgc cttaagttag cagtttcttc aagacttaga ccttttgtag ctcacacctg 2460
tacacaaatg ttgttatctg atatgtggat gggaaggctg aatatgagga aaaattcctg 2520
gtacaaggtc atactaagca ttttagttcc acctgccata ttgctgttag agtataaaac 2580
taaggctgaa atgtcccata tcccacaatc tcaagatgct catcagatga caatggatga 2640
cagcgaaaac aactttcaga acataacaga agagatcccc atggaagtgt ttaaagaagt 2700
acggattttg gatagtaatg aaggaaagaa tgagatggag atacaaatga aatcaaaaaa 2760
gcttccaatt acgcgaaagt tttatgcctt ttatcatgca ccaattgtaa aattctggtt 2820
taacacgttg gcatatttag gatttctgat gctttataca tttgtggttc ttgtacaaat 2880
ggaacagtta ccttcagttc aagaatggat tgttattgct tatattttta cttatgccat 2940
tgagaaagtc cgtgagatct ttatgtctga agctgggaaa gtaaaccaga agattaaagt 3000
atggtttagt gattacttca acatcagtga tacaattgcc ataatttctt tcttcattgg 3060
atttggacta agatttggag caaaatggaa ctttgcaaat gcatatgata atcatgtttt 3120
tgtggctgga agattaattt actgtcttaa cataatattt tggtatgtgc gtttgctaga 3180
ttttctagct gtaaatcaac aggcaggacc ttatgtaatg atgattggaa aaatggtggc 3240
caatatgttc tacattgtag tgattatggc tcttgtatta cttagttttg gtgttcccag 3300
aaaggcaata ctttatcctc atgaagcacc atcttggact cttgctaaag atatagtttt 3360
tcacccatac tggatgattt ttggtgaagt ttatgcatac gaaattgatg tgtgtgcaaa 3420
tgattctgtt atccctcaaa tctgtggtcc tgggacgtgg ttgactccat ttcttcaagc 3480
agtctacctc tttgtacagt atatcattat ggttaatctt cttattgcat tttcaacaa 3540
tgtgtattta caagtgaagg caatttccaa tattgtatgg aagtaccagc gttatcattt 3600
tattatggct tatcatgaga aaccagttct gcctcctcca cttatcattc ttagccatat 3660
agtttctctg ttttgctgca tatgtaagag aagaaagaaa gataagactt ccgatggacc 3720
aaaacttttc ttaacagaag aagatcaaaa gaaacttcat gattttgaag agcagtgtgt 3780
tgaaatgtat ttcaatgaaa aagatgacaa atttcattct gggagtgaag agagaattcg 3840
tgtcactttt gaaagagtgg aacagatgtg cattcagatt aaagaagttg gagatcgtgt 3900
caactacata aaaagatcat tacaatcatt agattctcaa attggccatt tgcaagatct 3960
ttcagccctg acggtagata cattaaaaac actcactgcc cagaaagcgt cggaagctag 4020
caaagttcat aatgaaatca cacgagaact gagcatttcc aaacacttgg ctcaaaacct 4080
tattgatgat ggtcctgtaa gaccttctgt atggaaaaag catggtgttg taaatacact 4140
tagctcctct cttcctcaag gtgatcttga aagtaataat ccttttcatt gtaatatttt 4200
aatgaaagat gacaaagatc cccagtgtaa tatatttggt caagacttac ctgcagtacc 4260
ccagagaaaa gaatttaatt ttccagaggc tggttcctct tctggtgcct tattcccaag 4320
tgctgtttcc cctccagaac tgcgacagag actacatggg gtagaactct taaaaatatt 4380
taataaaaat caaaaattag gcagttcatc tactagcata ccacatctgt catccccacc 4440
aaccaaattt tttgttagta caccatctca gccaagttgc aaaagccact tggaaactgg 4500
aaccaaagat caagaaactg tttgctctaa agctacagaa ggagataata cagaatttgg 4560
agcatttgta ggacacagag atagcatgga tttacagagg tttaaagaaa catcaaacaa 4620
gataaaaata ctatccaata acaatacttc tgaaaacact ttgaaacgag tgagttctct 4680
tgctggattt actgactgtc acagaacttc cattcctgtt cattcaaaac aagaaaaaat 4740
cagtagaagg ccatctaccg aagacactca tgaagtagat tccaaagcag ctttaatacc 4800
ggattggtta caagatagac catcaaacag agaaatgcca tctgaagaag gaacattaaa 4860
tggtctcact tctccattta agccagctat ggatacaaat tactattatt cagctgtgga 4920
aagaaataac ttgatgaggt tatcacagag cattccattt acacctgtgc ctccaagagg 4980
ggagcctgtc acagtgtatc gtttggaaga gagttcaccc aacatactaa ataacagcat 5040
gtcttcttgg tcacaactag gcctctgtgc caaaatagag tttttaagca aagaggagat 5100
gggaggaggt ttacgaagag ctgtcaaagt acagtgtacc tggtcagaac atgatatcct 5160
caaatcaggg catctttata ttatcaaatc ttttcttcca gaggtggtta atacatggtc 5220
aagtatttat aaagaagata cagttctgca tctctgtctg agagaaattc aacaacagag 5280
agcagcacaa aagcttacgt ttgcctttaa tcaaatgaaa cccaaatcca taccatattc 5340
tccaaggttc cttgaagttt tcctgctgta ttgccattca gcaggacagt ggtttgctgt 5400
ggaagaatgt atgactggag aatttagaaa atacaacaat aataatggag atgagattat 5460
tccaactaat actctggaag agatcatgct agcctttagc cactggactt acgaatatac 5520
aagaggggag ttactggtac ttgatttgca aggtgttggt gaaaatttga ctgacccatc 5580
tgtgataaaa gcagaagaaa agagatcctg tgatatggtt tttggcccag caaatctagg 5640
agaagatgca attaaaaact tcagagcaaa acatcactgt aattcttgct gtagaaagct 5700
taaacttcca gatctgaaga ggaatgatta tacgcctgat aaaattatat ttcctcagga 5760
```

tgagccttca gatttgaatc ttcagcctgg aaattccacc aaagaatcag aatcaactaa 5820
ttctgttcgt ctgatgttat aatattaata ttactgaatc attggttttg cctgcacctc 5880
acagaaatgt tactgtgtca cttttccctc gggaggaaat tgtttggtaa tatagaaagg 5940
tgtatgcaag ttgaatttgc tgactccagc acagttaaaa ggtcaatatt cttttgacct 6000
gattaatcag tcagaaagtc cctataggat agagctggca gctgagaaat tttaaaggta 6060
attgataatt agtatttata actttttaaa gggctctttg tatagcagag gatctcattt 6120
gactttgttt tgatgaggt gatgctctct cttatgtggt acaataccat taaccaaagg 6180
taggtgtcca tgcagatttt attggcagct g 6211

<210> 54
<211> 3714
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 6176128CB1

<400> 54
atggcgcggg ccaagctgcc gcgctcgccg tccgagggca aggcgggccc ggggggcgcc 60
ccagccggcg ccgcagcccc cgaggagcct cacgggctca gcccgctgct gccggcccgc 120
ggcgggggct ccgtgggcag cgacgtgggc cagaggcttc ctgtagaaga tttcagcctg 180
gactcctccc tgtctcaggt ccaggtggag ttctacgtca acgagaacac cttcaaggag 240
cggctcaagc tgttcttcat caaaaaccaa agatcgagtc tgaggatccg gctgttcaac 300
ttctccctga agctgctcac ctgcctgctc tacattgtgc gcgtcctgct cgatgacccg 360
gccctgggca tcggatggtg ggctgcccа aggcagaact actccttcaa tgactcgtcc 420
tccgagatca actgggctcc tattctgtgg gtggagagaa agatgacact gtgggcgatc 480
caggtcatcg tggccataat aagcttcctg gagacgatgc ttctcatcta cctcagctac 540
aaaggcaaca tctgggagca gatcttccgc gtgtccttcg tcctggagat gatcaacact 600
ctgcccttca tcatcacgat cttctggccg ccgctgcgga acctgttcat ccccgtcttt 660
ctgaactgct ggctggccaa gcacgcgctg gaaaacatga ttaatgactt ccaccgtgcc 720
atcctgcgga cacagtcagc catgttcaac caggtcctca tcctcttctg caccctgctg 780
tgcctcgttt tcacggggac ctgcggcatc cagcacctgg agcgggcggg cgagaacctg 840
tccctcctga cctccttcta cttctgcatc gtcaccttct ccaccgtggg ctacggtgac 900
gtcacgccca agatctggcc atcgcagctg ctggtggtca tcatgatctg cgtggccctc 960
gtggtgctcc cactgcagtt cgaggagctc gtctacctct ggatggagcg gcagaagtca 1020
gggggcaact acagccgcca ccgtgcgcag acggagaagc acgtggtcct gtgtgtcagc 1080
tccctcaaga tcgaccttct catggacttc ctgaacgagt tctacgccca cccccggctc 1140
caggactatt acgtggtcat cctgtgcccc acggagatgg atgtccaggt gcgcagagtc 1200
ctgcagatcc ctctgtggtc ccagcgggtc atctacctcc agggctctgc actcaaagac 1260
caggacctca tgcgagccaa gatggacaat ggggaggcct gcttcatcct cagcagcagg 1320
aacgaggtgg accgcacggc tgcagaccac cagaccatcc tgcgcgcctg ggccgtgaag 1380
gacttcgccc ccaactgccc cctctacgtc cagatcctca aacctgaaaa caagtttcac 1440
gtcaagtttg ctgaccacgt ggtgtgtgag gaggagtgca agtacgccat gctggcgctg 1500
aactgcatct gcccggcgac ctccaccctc atcaccctgc tggtgcacac gtcccgcggc 1560
caggagggac aggagtctcc ggagcagtgg cagcgcatgt atgggcgctg ctccggcaac 1620
gaggtgtacc acatccgcat gggtgacagc aagttcttcc gcgagtacga gggcaagagc 1680
ttcacctacg cggccttcca cgcccacaag aagtatggcg tgtgcctcat cggcctgaag 1740
cgggaggaca acaagagcat cctgctgaac ccggggcccc ggcacatcct ggccgcctct 1800
gacacctgct tctacatcaa catcaccaag gaggagaact cggccttcat cttcaagcag 1860
gaggagaagc ggaagaagag ggccttctcg gggcaggggc tgcacgaggg tccggcccgc 1920
ctgcccgtgc acagcatcat cgcctccatg gggacagtgg ccatggacct gcagggcaca 1980
gagcaccggc ctacgcagag cggcggtggg ggcgggggca gcaagctggc actgcccacg 2040
gagaacggct cgggcagccg gcggcccagc atcgcgcccg tcctggaact ggccgacagc 2100
tcagccctgc tgccctgcga cctgctgagc gaccagtcgg aggatgaggt gacgccgtcg 2160
gacgacgagg ggctctccgt ggtagagtat gtgaagggct accctcccaa ctcgccctac 2220
atcggcagct ccccaaccct gtgccacctc ctgcctgtga aagccccctt ctgctgcctg 2280
cggctggaca agggctgcaa gcacaacagc tatgaagacg ccaaggccta cgggttcaag 2340
aacaagctga tcatcgtctc ggcagagacg gccggcaatg ggctgtacaa cttcatcgtg 2400
ccactgcggg cctactacag atccccgcaag gagctgaacc ccatcgtgct gctgctggac 2460
aacaagcccg accaccactt cctggaagcc atctgctgct tccccatggt ctactacatg 2520
gagggctctg tggacaacct ggacagcctg ctgcagtgtg gcatcatcta tgcggacaac 2580
ctggtggtgg tggacaagga gagcaccatg agcgccgagg aggactacat ggcggacgcc 2640

aagaccatcg tcaacgtgca gaccatgttc cggctcttcc ccagcctcag catcaccacg 2700
gagctcaccc acccttccaa catgcgcttc atgcagttcc gcgccaagga cagctactct 2760
ctggctcttt ccaaactaga aaagagggag cgagagaatg gctccaacct ggccttcatg 2820
ttccgcctgc cgttcgccgc cggccgcgtc ttcagcatca gcatgttgga cacactgctc 2880
taccagtcct tcgtgaagga ctacatgatc accatcaccc ggctgctgct gggcctggac 2940
accacgccgg gctcggggta cctctgtgcc atgaaaatca ccgagggcga cctgtggatc 3000
cgcacgtacg gccgcctctt ccagaagctc tgctcctcca gcgccgagat ccccattggc 3060
atctaccgga cagagagcca cgtcttctcc acctcggagc cccacgaact cagagcccag 3120
tcccagatct cggtgaacgt ggaggactgt gaggacacac gggaagtgaa ggggccctgg 3180
ggctcccgcg ctggcaccgg aggcagctcc cagggccgcc acacgggcgg cggtgacccc 3240
gcagagcacc cactgctacg gcgcaagagc ctgcagtggg cccggaggct gagccgcaag 3300
gcgcccaagc aggcaggccg ggcggcggcc gcggagtgga tcagccagca gcgcctcagc 3360
ctgtaccggc gctctgagcg ccaggagctc tccgagctgg tgaagaaccg catgaagcac 3420
ctggggctgc ccaccaccgg ctacgaggac gtagcaaatt taacagccag tgatgtcatg 3480
aatcgggtaa acctgggata tttgcaagac gagatgaacg accaccagaa caccctctcc 3540
tacgtcctca tcaaccctcc gcccgacacg aggctggagc ccagtgacat tgtgtatctc 3600
atccgctccg accccctggc tcacgtggcc agcagctccc agagccggaa gagcagctgc 3660
agccacaagc tgtcgtcctg caaccccgag actcgcgacg agacacagct ctga 3714

<210> 55
<211> 3115
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7473418CB1

<400> 55
atggcctcgg cgctgagcta tgtctccaag ttcaagtcct tcgtgatctt gttcgtcacc 60
ccgctcctgc tgctgccact cgtcattctg atgcccgcca agtttgtcag gtgtgcctac 120
gtcatcatcc tcatggccat ttactggtgc acagaagtca tccctctggc tgtcacctct 180
ctcatgcctg tcttgctttt cccactcttc cagattctgg actccaggca ggtgtgtgtc 240
cagtacatga aggacaccaa catgctgttc ctgggcggcc tcatcgtggc cgtggctgtg 300
gagcgctgga acctgcacaa gaggatcgcc ctgcgcacgc tcctctgggt gggggccaag 360
cctgcacggc tgatgctggg cttcatgggc gtcacagccc tcctgtccat gtggatcagt 420
aacacggcaa ccacggccat gatggtgccc atcgtggagg ccatattgca gcagatggaa 480
gccacaagcg cagccaccga ggccggcctg gagctggtgg acaagggcaa ggccaaggag 540
ctgccagcta acagcgctgt gcccaccaca gggagtcaag tgatttttga aggcccccact 600
ctggggcagc aggaagacca agagcggaag aggttgtgta aggccatgac cctgtgcatc 660
tgctacgcgg ccagcatcgg gggcaccgcc accctgaccg ggacgggacc caacgtggtg 720
ctcctgggcc agatgaacga gttgtttcct gacagcaagg acctcgtgaa ctttgcttcc 780
tggtttgcat ttgcctttcc caacatgctg gtgatgctgc tgttcgcctg gctgtggctc 840
cagtttgttt acatgagatt caattttaaa aagtcctggg gctgcgggct agagagcaag 900
aaaaacgaga aggctgccct caaggtgctg caggaggagt accggaagct ggggcccttg 960
tccttcgcgg agatcaacgt gctgatctgc ttcttcctgc tggtcatcct gtggttctcc 1020
cgagaccccg gcttcatgcc cggctggctg actgttgcct gggtggagga aaggaaaact 1080
ccatttatc cccctcccct gctggattgg aaggtaaccc aggagaaagt gccctggggc 1140
atcgtgctgc tactaggggg cggatttgct ctggctaaag gatccgaggc ctcggggctg 1200
tccgtgtgga tggggaagca gatggagccc ttgcacgcag tgccccggc agccatcacc 1260
ttgatcttgt ccttgctcgt tgccgtgttc actgagtgca caagcaacgt ggccaccacc 1320
accttgttcc tgcccatctt tgcctccatg tctcgctcca tcggcctcaa tccgctgtac 1380
atcatgctgc cctgtaccct gagtgcctcc tttgccttca tgttgcctgt ggccacccct 1440
ccaaatgcca tcgtgttcac ctatgggcac ctcaaggttg ctgacatggt gaaaacagga 1500
gtcataatga acataattgg agtcttctgt gtgtttttgg ctgtcaacac ctggggacgg 1560
gccatatttg acttggatca tttccctgac tgggctaatg tgacacatat tgagacttag 1620
gaagagccac aagaccacac acacagccct taccctcctc aggactaccg aaccttctgg 1680
cacaccttgt acagagtttt ggggttcaca ccccaaaatg acccaacgat gtccacacac 1740
caccaaaacc cagccaatgg gccacctctt cctccaagcc cagatgcaga gatggacatg 1800
ggcagctgga gggtaggctc agaaatgaag ggaacccctc agtgggctgc tggacccatc 1860
tttcccaagc cttgccatta tctctgtgag ggaggccagg tagccgaggg atcaggatgc 1920
aggctgctgt acccgctctg cctcaagcat cccccacaca gggctctggt tttcactcgc 1980
ttcgtcctag atagtttaaa tgggaatcgg atcccctggt tgagagctaa gacaaccacc 2040

taccagtgcc catgtccctt ccagctcacc ttgagcagcc tcagatcatc tctgtcactc 2100
tggaagggac accccagcca gggacggaat gcctggtctt gagcaacctc ccactgctgg 2160
agtgcgagtg ggaatcagag cctcctgaag cctctgggaa ctcctcctgt ggccaccacc 2220
aaaggatgag gaatctgagt tgccaacttc aggacgacac ctggcttgcc acccacagtg 2280
caccacaggc caacctacgc ccttcatcac ttggttctgt tttaatcgac tggccccctg 2340
tcccacctct ccagtgagcc tccttcaact ccttggtccc ctgttgtctg ggtcaacatt 2400
tgccgagacg ccttggctgg caccctctgg ggtccccctt ttctcccagg caggtcatct 2460
tttctgggag atgcttcccc tgccatcccc aaatagctag gatcacactc caagtatggg 2520
cagtgatggc gctctggggg ccacagtggg ctatctaggc cctccctcac ctgaggccca 2580
gagtggacac agctgttaat ttccactggc tatgccactt cagagtcttt catgccagcg 2640
tttgagctcc tctgggtaaa atcttccctt tgttgactgg ccttcacagc catggctggt 2700
gacaacagag gatcgttgag attgagcagc gcttggtgat ctctcagcaa acaacccctg 2760
cccgtgggcc aatctacttg aagttactcg gacaaagacc ccaaagtggg gcaacaactc 2820
cagagaggct gtgggaatct tcagaagccc ccctgtaaga gacagacatg agagacaagc 2880
atcttctttc ccccgcaagt ccattttatt tccttcttgt gctgctctgg aagagaggca 2940
gtagcaaaga gatgagctcc tggatggcat tttccagggc aggagaaagt atgagagcct 3000
caggaaaccc catcaaggac cgagtatgtg tctggttcct tgggtgggac gattcctgac 3060
cacactgtcc agctcttgct ctcattaaat gctctgtctc ccgcggaaaa aaaaa 3115

<210> 56
<211> 2846
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7474129CB1

<400> 56
ttccagccat ccctctgcct gcaatgagag cttcccgccg cctcagccac agtcccaccc 60
gggggccttg ggccccagac atgcggtgat ctcagggcaa gggttgccac gaccacccag 120
aacctcacca gccatgaaag cccaccccaa ggagatggtg cctctcatgg gcaagagagt 180
tgctgccccc agtgggaacc ctgccgtcct gccagagaag aggccggcgg agatcacccc 240
cacaaagaag agcatctctg gtaactgtga tgacatggac tccccccagt ctcctcaaga 300
tgatgtgaca gagaccccat ccaatcccaa cagccccagt gcacagctgg ccaaggaaga 360
gcagaggagg aaaaagaggc ggctgaagaa gcgcatcttt gcagccgtgt ctgagggctg 420
cgtggaggag ttggtagagt tgctggtgga gctgcaggag ctttgcaggc ggcgccatga 480
tgaggatgtg cctgacttcc tcatgcacaa gctgacggcc tccgacacgg ggaagacctg 540
cctgatgaag gccttgttaa acatcaaccc caacaccaag gagatcgtgc ggatcctgct 600
tgcctttgct gaagagaacg acatcctggg caggttcatc aacgccgagt acacagagga 660
ggcctatgaa gggcagacgg cgctgaacat cgccatcgag cggcggcagg gggacatcgc 720
agccctgctc atcgccgccg gcgccgacgt caacgcgcac gccaaggggg ccttcttcaa 780
ccccaagtac caacacgaag gcttctactt cggtgagacg cccctggccc tggcagcatg 840
caccaaccag cccgagattg tgcagctgct gatggagcac gagcagacgg acatcacctc 900
gcgggactca cgaggcaaca acatccttca cgccctggtg accgtggccg aggacttcaa 960
gacgcagaat gacgttgtga agcgcatgta cgacatgatc ctactgcgga gtggcaactg 1020
ggagctggag accactcgca acaacgatgg cctcacgccg ctgcagctgg ccgccaagat 1080
gggcaaggcg gagatcctga agtacatcct cagtcgtgag atcaaggaga agcggctccg 1140
gagcctgtcc aggaagttca ccgactgggc gtacggaccc gtgtcatcct ccctctacga 1200
cctcaccaac gtggacacca ccacggacaa ctcagtgctg gaaatcactg tctacaacac 1260
caacatcgac aaccggcatg agatgctgac cctggagccg ctgcacacgc tgctgcatat 1320
gaagtggaag aagtttgcca agcacatgtt ctttctgtcc ttctgctttt atttcttcta 1380
caacatcacc ctgaccctcg tctcgtacta ccgcccccgg gaggaggagg ccatcccgca 1440
ccccttggcc ctgacgcaca agatggggtg gctgcagctc ctagggagga tgtttgtgct 1500
catctgggcc atgtgcatct ctgtgaaaga gggcattgcc atcttcctgc tgagaccctc 1560
ggatctgcag tccatcctct cggatgcctg gttccacttt gtcttttttа tccaagctgt 1620
gcttgtgata ctgtctgtct tcttgtactt gtttgcctac aaagagtacc tcgcctgcct 1680
cgtgctggcc atggccctgg gctgggcgaa catgctctac tatacgcggg gtttccagtc 1740
catgggcatg tacagcgtca tgatccagaa ggtcattttg catgatgttc tgaagttctt 1800
gtttgtatat atcgtgtttt tgcttggatt tggagtagcc ttggcctcgc tgatcgagaa 1860
gtgtcccaaa gacaacaagg actgcagctc ctacggcagc ttcagcgacg cagtgctgga 1920
actcttcaag ctcaccatag gcctgggtga cctgaacatc cagcagaact ccaagtatcc 1980
cattctcttt ctgttcctgc tcatcaccta tgtcatcctc acctttgttc tcctcctcaa 2040

catgctcatt gctctgatgg gcgagactgt ggagaacgtc tccaaggaga gcgaacgcat 2100
ctggcgcctg cagagagcca ggaccatctt ggagtttgag aaaatgttac cagaatggct 2160
gaggagcaga ttccggatgg gagagctgtg caaagtggcc gaggatgatt tccgactgtg 2220
tttgcggatc aatgaggtga agtggactga atggaagacg cacgtctcct tccttaacga 2280
agacccgggg cctgtaagac gaacagattt caacaaaatc caagattctt ccaggaacaa 2340
cagcaaaacc actctcaatg catttgaaga agtcgaggaa ttcccggaaa cctcggtgta 2400
gaagcggaac ccagagctgg tgtgcgcgtg cgctgtctgg cgctgcaggc ggagtcaccg 2460
actctgtgca gagaggcttt gagggatgat ggagtccggc tctgctggcc tagaagcaga 2520
gtgcaccctc gtgctcagtg ctcagtgggt gtctgaactg aggggcagtt gtcaatttgt 2580
ctgagtggga aacatcctgg attttgttac ttggcaaaca gctggtgtaa acctacagcc 2640
agcagcagtc tggagcctgg gagcctcctg aagtcccggg tgaagcctct ggttttacca 2700
attgcaggtc ggcttggctg ggagagatgg atggcgggaa aggggcagca gtcttgagga 2760
gcagggagag gagtctttcc tcctgccagc ttcccccgtc agccccaacc ccagcccaca 2820
cattgtacca tctcttctgc tgtgac 2846

<210> 57
<211> 906
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7481414CB1

<400> 57
atgaagtctc accctgccat ccaagccgcc atagacctca ctgcgggcgc agcagggggc 60
ggagcttgtg tgctgactgg gcaacccttc gacaccataa aagtaaagat gcagacattt 120
cctcagctgt acaaaggcct tgccgactgc ttcctgaaaa catacaacca agtgggcatc 180
cgtggccttt acaggggaac cagtcctgca ctgctagcct atgtcaccca gggttctgtc 240
ctgttcatgt gctttggctt ttgccaacag tttgtcagga aagtggctag agtggagcag 300
aatgcagagc tgaacgactt ggagactgct actgctgggt cgctggcttc tgcatttgct 360
gcgctggctc tctgccccac tgagcttgtg aagtgtcggc tgcagaccat gtatgagatg 420
aagatgtcag ggaagatagc acaaagctat aacacaattt ggtctatggt taagagtatc 480
ttcatgaagg atggtccctt aggcttctat cgtggactct cgaccactct tgctcaggaa 540
atacctggct atttcttcta ctttggggc tatgaaatca gtcgatcatt ttttgcatca 600
gggggatcaa aggatgaact aggccctgtc cctttgatgt taagtggagg ctttgctggg 660
atctgtctct ggcttatcat attcccagtg gactgcatta aatccagaat ccaggttctt 720
tctatgtttg ggaagcctgc aggattaatc gaaacctta taagtgttgt gagaaatgaa 780
gggatatcag ccttgtattc tggattgaaa gccactctga ttcgagccat cccttccaat 840
gctgctctct tttggtttta tgagtacagc agaaagatga tgatgaacat ggtggaagaa 900
tactga 906

<210> 58
<211> 1840
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7481461CB1

<400> 58
gcggccgcct gcgcgctggc cgcctgcgcg ctgccagccc gcccgcccgc caggggctcc 60
gccgccctcg cctcggcctc gttagcccgc caggagcccc gcagctcctc cgggagcccg 120
ctggtaactc gcgtccctcg cgcttctccg gcgcctgagg ggcccgcctc gggccatggt 180
gctctcccag gaggagccgg actccgcgcg gggcacgagc gaggcgcagc cgctcggccc 240
cgcgcccacg ggggccgctc cgccgcccgg cccgggaccc tcggacagcc ccgaggcggc 300
tgtcgagaag gtggaggtgg agctggcggg gccggcgacc gcggagcccc atgagccccc 360
cgaacccccc gagggcggct ggggctggct ggtgatgctg gcggccatgt ggtgcaacgg 420
gtcggtgttc ggcatccaga acgcttgcgg ggtgctcttc gtgtccatgc tggaaacctt 480
cggctccaaa gacgatgaca agatggtctt taagacagca tgggtaggtt ctctctccat 540
ggggatgatt ttcttttgct gcccaatagt cagcgtcttc acagacctat ttggttgtcg 600
gaaaacagct gtcgtgggtg ctgctgttgg atttgttggg ctcatgtcca gttcttttgt 660

aagttccatc gagcctctgt accttaccta tggaatcata tttgcctgcg gctgctcctt 720
tgcataccag ccttcattgg tcattttggg acactatttc aagaagcgcc ttggactggt 780
gaatggcatt gtcactgctg gcagcagtgt cttcacaatc ctgctgcctt tgctcttaag 840
ggttctgatt gacagcgtgg gcctctttta cacattgagg gtgctctgca tcttcatgtt 900
tgttctcttt ctggctggct ttacttaccg acctcttgct accagtacca aagataaaga 960
gagtggaggt agcggatcct ccctcttttc caggaaaaag ttcagtcctc caaaaaaaat 1020
tttcaatttt gccatcttca aggtgacagc ttatgcagtg tgggcagttg gaataccact 1080
tgcacttttt ggatactttg tgccttatgt tcacttgatg aaacatgtaa atgaaagatt 1140
tcaagatgaa aaaaataaag aggttgttct catgtgcatt ggcgtcactt caggagttgg 1200
acgactgctc tttggccgga ttgcagatta tgtgcctggt gtgaagaagg tttatctaca 1260
ggtactctcc tttttcttca ttggtctgat gtccatgatg attcctctgt gtagcatctt 1320
tggggccctc attgctgtgt gcctcatcat gggtctcttc gatggatgct tcatttccat 1380
tatggctccc atagcctttg agttagttgg tgcccaggat gtctcccaag caattggatt 1440
tctgctcgga ttcatgtcta tacccatgac tgttggccca cccattgcag ggttacttcg 1500
tgacaaactg ggctcctatg atgtggcatt ctacctcgct ggagtccctc cccttattgg 1560
aggtgctgtg ctttgtttta tcccgtggat ccatagtaag aagcaaagag agatcagtaa 1620
aaccactgga aaagaaaaga tggagaaaat gttggaaaac cagaactctc tgctgtcaag 1680
ttcatctgga atgttcaaga aagaatctga ctctattatt taatatctta catacctcca 1740
ccagactgga cttgcttttt gaattttaag caagtttcct ttccttttat acaaattgca 1800
aatttcatat ttttttaatc acatcctagg aatagcacaa 1840

<210> 59
<211> 5348
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7472541CB1

<400> 59
gttagcttga ggccttgcct tgacataatc agatataatc agaaaaatga gaaattccat 60
aggaaagaga actattttag ccaaggtgtg cgagagaaat accgccactt tcaagcactg 120
ttttcttcta ctggagtctg ctcaataggg acgtcagctt tgctggggct tcctttgaca 180
agagaatcag aaccgactgg tgacatttgt ttcaatgaaa gcaacagtgt gaagaaactt 240
gcaatttatt ctgatcaaca atttgcccag ctaaagtact acatctgccc cctcttcctg 300
tggctgtagg ggcacagcaa aggtcactgg tctaacctcc ttaaagggac tccgctaaca 360
gatcttcgcc tgctgctgga aatggccctc tcagtggact catcgtggca tcggtggcag 420
tggagagtca gagatggctt cccccattgt ccatcggaaa ccacaccgct gctctctcca 480
gagaaaggga gacagagcta caacttgaca cagcagcggg tcgtgttccc caacaacagc 540
atattccatc aagattggga agaggtctcc aggagatacc ctggcaacag aacctgcaca 600
accaaataca ccctcttcac cttcctgccc cggaatctct ttgagcaatt tcatagatgg 660
gctaacctct atttcctgtt cctggtgatt ttgagctgga tgccctccat ggaagtcttc 720
cacagagaaa tcaccatgtt accattggcc attgtcctgt tcgtcatcat gatcaaggat 780
ggcatggagg acttcaagag acaccgcttt gataaagcaa taaactgctc caacattcga 840
atttatgaaa gaaaagagca gacctatgtg cagaagtgct ggaaggatgt gcgtgtggga 900
gacttcatcc aaatgaaatg caatgagatt gtcccagcag acatactcct ccttttttcc 960
tctgacccca atgggatatg ccatctggaa actgccagct tggatggaga gacaaacctc 1020
aagcaaagac gtgtcgtgaa gggcttctca cagcaggagg tacagttcga accagagctt 1080
ttccacaata ccatcgtgtg tgagaaaccc aacaaccacc tcaacaaatt taagggttat 1140
atggagcatc ctgaccagac caggactggc tttggctgtg agagtcttct gcttcgaggc 1200
tgcaccatca gaaacaccga gatggctgtt ggcattgtca tctatgcagg ccatgagacg 1260
aaagccatgc tgaacaacag tggcccccgg tacaaacgca gcaagattga gcggcgcatg 1320
aatatagaca tcttcttctg cattgggatc ctcatcctca tgtgccttat tggagctgta 1380
ggtcacagca tctggaatgg gacctttgaa gaacaccctc ccttcgatgt gccagatgcc 1440
aatggcagct tccttcccag tgcccttggg ggcttctaca tgttcctcac aatgatcatc 1500
ctgctccagg tgctgatccc catctctttg tatgtctcca ttgagctggt gaagctcggg 1560
caagtgttct tcttgagcaa tgaccttgac ctgtatgatg aagagaccga tttatccatt 1620
caatgtcgag ccctcaacat cgcagaggac ttgggccaga tccagtacat cttctccgat 1680
aagacgggga ccctgacaga gaacaagatg gtgttccgac gttgcaccat catgggcagc 1740
gagtattctc accaagaaaa tgctaagcga ctggagaccc caaaggagct ggactcagat 1800
ggtgaagagt ggacccaata ccaatgcctg tccttctcgg ctagatgggc ccaggatcca 1860
gcaactatga gaagccaaaa aggtgctcag cctctgagga ggagccagag tgcccgggtg 1920

cccatccagg gccactaccg gcaaaggtct atggggcacc gtgaaagctc acagcctcct 1980
gtggccttca gcagctccat agaaaaagat gtaactccag ataaaaacct actgaccaag 2040
gttcgagatg ctgccctgtg gttggagacc ttgtcagaca gcagacctgc caaggcttcc 2100
ctctccacca cctcctccat tgctgatttc ttccttgcct taaccatctg caactctgtc 2160
atggtgtcca caaccaccga gcccaggcag aggtgggatg atcaaaagat agtggaaaat 2220
gaccattgtc aatgcttaga atttcagggc tggaggaaaa tatctggctt cacttattgc 2280
aaaagtacct tcatattccg cataagacaa cttggtatta tttccaacat tgagagtaat 2340
attccacttt ccttctttgg ccacaaggtc accatcaaac cctcaagcaa ggctctgggg 2400
acgtccctgg agaagattca gcagctcttc cagaagttga agctattgag cctcagccag 2460
tcattctcat ccactgcacc ctctgacaca gacctcgggg agagcttagg ggccaacgtg 2520
gccaccacag actcggatga gagagatgat gcatctgtgt gcagtggagg tgactccact 2580
gatgacggtg gctacaggag cagcatgtgg gaccagggcg acatcctgga gtctgggtca 2640
ggcacttcct tggaggaggc attggaggcc ccagccacag acctggccag gcctgagttc 2700
tgttacgagg ctgagagccc tgatgaggcc gccctggtgc acgctgccca tgcctacagc 2760
ttcacactag tgtcccggac acctgagcag gtgactgtgc gcctgcccca gggcacctgc 2820
ctcaccttca gcctcctctg caccctgggc tttgactctg tcaggaagag aatgtctgtg 2880
gttgtgaggc acccactgac tggcgagatt gttgtctaca ccaagggtgc tgactcggtc 2940
atcatggacc tgctggaaga cccagcctgc gtacctgaca ttaatatgga aaagaagctg 3000
agaaaaatcc gagcccggac ccaaaagcat ctagacttgt atgcaagaga tggcctgcgc 3060
acactatgca ttgccaagaa ggttgtaagc gaagaggact tccggagatg ggccagtttc 3120
cggcgtgagg ctgaggcatc cctcgacaac cgagatgagc ttctcatgga aactgcacag 3180
catctggaga atcaactcac cttacttgga gccactggga tcgaagaccg gctgcaggaa 3240
ggagttccag atacgattgc cactctgcgg gaggctggga tccagctctg ggtcctgact 3300
ggagataagc aggagacagc ggtcaacatt gcccattcct gcagactgtt aaatcagacc 3360
gacactgttt ataccatcaa tacagagaat caggagacct gtgaatccat cctcaattgt 3420
gcattggaag agctaaagca atttcgtgaa ctacagaagc cagaccgcaa gctctttgga 3480
ttccgcttac cttccaagac accatccatc acctcagaag ctgtggttcc agaagctgga 3540
ttggtcatcg atgggaagac attgaatgcc atcttccagg gaaagctaga gaagaagttt 3600
ctggaattga cccagtattg tcggtccgtc ctgtgctgcc gctccacgcc actccagaag 3660
agtatgatag tcaagctggt gcgagacaag ttgcgcgtca tgacccttc catagtgat 3720
ggagcaaatg atgtaagcat gattcaagct gctgatattg gaattggaat atctggacag 3780
gaaggcatgc aggctgtcat gtccagcgac tttgccatca cccgctttaa gcatctcaag 3840
aagttgctgc tcgtgcatgg ccactggtgt tactcgcgcc tggccaggat ggtggtgtac 3900
tacctctaca agaacgtgtg ctacgtcaac ctgctcttct ggtatcagtt cttctgtggt 3960
ttctccagct ccaccatgat tgattactgg cagatgatat tcttcaatct cttctttacc 4020
tccttgcctc ctcttgtctt tggagtcctt gacaaagaca tctctgcaga aacactcctg 4080
gcattgcctg agctatacaa gagtggccag aactctgagt gctataacct gtcgactttc 4140
tggatttcta tggtggatgc attctaccag agcctcatct gtttctttat cccttacctg 4200
gcctataagg gctctgatat agatgtcttt acctttggga caccaatcaa caccatctcc 4260
ctcaccacaa tccttttgca ccaggcaatg gaaatgaaga catggaccat tttccacgga 4320
gtcgtgctcc tcggcagctt cctgatgtac tttctggtat ccctcctgta caatgccacc 4380
tgcgtcatct gcaacagccc caccaatccc tattgggtga tggaaggcca gctctcaaac 4440
cccactttct acctcgtctg ctttctcaca ccagttgttg ctcttctccc aagatacttt 4500
ttcctgtctc tgcaaggaac ttgtgggaag tctctaatct caaaagctca gaaaattgac 4560
aaactccccc cagacaaaag aaacctggaa atccagagtt ggagaagcag acagaggcct 4620
gcccctgtcc ccgaagtggc tcgaccaact caccacccag tgtcatctat cacaggacag 4680
gacttcagtg ccagcacccc aaagagctct aaccctccca agaggaagca tgtggaagag 4740
tcagtgctcc acgaacagag atgtggcacg gagtgcatga gggatgactc atgctcaggg 4800
gactcctcag ctcaactctc atccggggag cacctgctgg gacctaacag gataatggcc 4860
tactcaggag gacagactga tatgtgccgg tgctcaaaga ggagcagcca tcgccgatcc 4920
cagagttcac tgaccatatg aggagctgca gaaatctgta caaactcaac agaggccacc 4980
tagtcactgg tccacataac ccttgacccc ttcttcttca tagaggaaac aatgtgccag 5040
tcttattctt ttcttcaaca accttgactt ccatggagga agtgctggcc ccaaggggtc 5100
tgacacaaag acgggaaacc cagtcggcct ctagttttct gctgctctca ggcagcacat 5160
cttgcaaaca gtttggagaa ggaggctgtt tttgttgaat cgagttctca aatcggttta 5220
gaccaaagcc attcttctga ccctctagat aagcgtagcc tacaacccag tgccgtaagt 5280
ttccaagatt caagaagtgt atcaacccag gcaatatctc aggatatgga agtttctggg 5340
tttattta 5348

<210> 60
<211> 5149
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 6999183CB1

<400> 60
atgagcaaga gacgcatgag cgtgggtcag caaacatggg ctcttctctg caagaactgt 60
ctcaaaaaat ggagaatgaa aagacagacc ttgttggaat ggctcttttc atttcttctg 120
gtactgtttc tgtacctatt tttctccaat ttacatcaag ttcatgacac tcctcaaatg 180
tcttcaatgg atctgggacg tgtagatagt tttaatgata ctaattatgt tattgcattt 240
gcacctgaat ccaaaactac ccaagagata atgaacaaag tggcttcagc cccattccta 300
atggcaggaa gaacaatcat ggggtggcct gatgaaaaaa gcatggatga attggatttg 360
aactattcaa tagacgcagt gagagtcatc tttactgata ccttctccta ccatttgaag 420
ttttcttggg gacatagaat ccccatgatg aaagagcaca gagaccattc agctcactgt 480
caagcagtga atgaaaaaat gaagtgtgaa ggttcagagt tctgggagaa aggctttgta 540
gcttttcaag ctgccattaa tgctgctatc atagaaatcg caacaaatca ttcagtgatg 600
gaacagctga tgtcagttac tggtgtacat atgaagatat taccttttgt tgcccaagga 660
ggagttgcaa ctgatttttt cattttcttt tgcattattt ctttttctac atttatatac 720
tatgtatcag tcaatgttac acaagaaaga caatacatta cgtcattgat gacaatgatg 780
ggactccgag agtcagcatt ctggctttcc tggggtttga tgtatgctgg cttcatcctt 840
atcatggcca ctttaatggc tcttattgta aaatctgcac aaattgtcgt cctgactggt 900
tttgtgatgg tcttcaccct ctttctcctc tatggcctgt ctttgataac tttagctttc 960
ctgatgagtg tgttgataaa gaaacctttc cttacgggct tggttgtgtt tctccttatt 1020
gtctttgggg ggatcctggg attcccagca ttgtatacac atcttcctgc atttttggaa 1080
tggactttgt gtcttcttag ccccttttgcc ttcactgttg ggatggccca gcttatacat 1140
ttggactatg atgtgaattc taatgcccac ttggattctt cacaaaatcc atacctcata 1200
atagctactc ttttcatgtt ggtttttgac acccttctgt atttggtatt gacattatat 1260
tttgacaaaa ttttgcccgc tgaatatgga catcgatgtt ctcccttgtt tttcctgaaa 1320
tcctgttttt ggtttcaaca cggaagggct aatcatgtgg tccttgagaa tgaaacagat 1380
tctgatccta cacctaatga ctgttttgaa ccagtgtctc cagaattctg tgggaaggaa 1440
gccatcagaa tcaaaaatct taaaaaagaa tatgcaggga agtgtgagag agtagaagct 1500
ttgaaaggtg tggtgtttga catatatgaa ggccagatca ctgccctcct tggtcacagt 1560
ggagctggaa aaactaccct gttaaacata cttagtgggt tgtcagttcc aacatcaggt 1620
tcagtcactg tctataatca cacactttca agaatggctg atatagaaaa tatcagcaag 1680
ttcactggat tttgtccaca atccaatgtg caatttggat ttctcactgt gaaagaaaac 1740
ctcaggctgt ttgctaaaat aaaagggatt ttgccacatg aagtggagaa agaggtattg 1800
ctattggatg aaccgactgc tggattggat cctctttcaa ggcaccgaat atggaatctc 1860
ctgaaagagg ggaaatcaga cagagtaatt ctcttcagca cccagtttat agatgaggct 1920
gacattctgg cggacaggaa ggtgttcata tccaatggga agctgaagtg tgcaggctct 1980
tctctgttcc ttaagaagaa atggggcata ggctaccatt taagtttgca tctgaatgaa 2040
aggtgtgatc cagagagtat aacatcactg gttaagcagc acatctctga tgccaaattg 2100
acagcacaaa gtgaagaaaa acttgtatat attttgcctt tggaaaggac aaacaaattt 2160
ccagaacttt acagggatct tgatagatgt tctaaccaag gcattgagga ttatggtgtt 2220
tccataacaa ctttgaatga ggtgtttctg aaattagaag gaaaatcaac tattgatgaa 2280
tcagatattg gaatttgggg acaattacaa actgatgggg caaaagatat aggaagcctt 2340
gttgagctgg aacaagtttt gtcttccttc cacgaaacaa ggaaaacaat cagtggcgtg 2400
gcgctctgga ggcagcaggt ctgtgcaata gcaaaagttc gcttcctaaa gttaaagaaa 2460
gaaagaaaaa gcctgtggac tatattattg ctttttggta ttagctttat ccctcaactt 2520
ttggaacatc tattctacga gtcatatcag aaaagttacc cgtgggaact gtctccaaat 2580
acatacttcc tctcaccagg acaacaacca caggatcctc tgacccattt actggtcatc 2640
aataagacag ggtcaaccat tgataacttt ttacattcac tgaggcgaca gaacatagct 2700
atagaagtgg atgcctttgg aactagaaat ggcacagatg acccatctta caatggtgct 2760
atcattgtgt caggtgatga aaaggatcac agattttcaa tagcatgtaa tacaaaacgg 2820
ctgaattgct ttcctgtcct cctggatgtc attagcaatg gactacttgg aatttttaat 2880
tcgtcagaac acattcagac tgacagaagc acatttttg aagagcatat ggattatgag 2940
tatgggtacc gaagtaacac cttcttctgg ataccgatgg cagcctcttt cactccatac 3000
attgcaatga gcagcattgg tgactacaaa aaaaaagctc attcccagct acggatttca 3060
ggcctctacc cttctgcata ctggtttggc caagcactgg tggatgtttc cctgtacttt 3120
ttgatcctcc tgctaatgca aataatggat tatattttta gcccagagga gattatattt 3180
ataattcaaa acctgttaat tcaaatcctg tgtagtattg gctatgtctc atctcttgtt 3240
ttcttgacat atgtgatttc attcattttt cgcaatggga gaaaaaatag tggcatttgg 3300
tcatttttct tcttaattgt ggtcatcttc tcgatagttg ctactgatct aaatgaatat 3360
ggatttctag ggctattttt tggcaccatg ttaatacctc ccttcacatt gattggctct 3420
ctattcattt tttctgagat ttctcctgat tccatggatt acttaggagc ttcagaatct 3480

```
gaaattgtat acctggcact gctaatacct taccttcatt ttctcatttt tcttttcatt 3540
ctgcgatgcc tagaaatgaa ctgcaggaag aaactaatga gaaaggatcc tgtgttcaga 3600
atttctccaa gaagcaacgc tatttttcca aacccagaag agcctgaagg agaggaggaa 3660
gatatccaga tggaaagaat gagaacagtg aatgctatgg ctgtgcgaga ctttgatgag 3720
acacccgtca tcattgccag ctgtctacgg aaggaatatg caggcaaaaa gaaaaattgc 3780
ttttctaaaa ggaagaaaac aattgccaca agaaatgtct ctttttgtgt taaaaaaggt 3840
gaagttatag gactgttagg acacaatgga gctggtaaaa gtacaactat taagatgata 3900
actggagaca caaaaccaac tgcaggacag gtgattttga aagggagcgg tggaggggaa 3960
cccctgggct tcctggggta ctgccctcag gagaatgcgc tgtggcccaa cctgacagtg 4020
aggcagcacc tggaggtgta cgctgccgtg aaaggtctca ggaaaggggga cgcaatgatc 4080
gccatcacac ggttagtgga tgcgctcaag ctgcaggacc agctgaaggc tcccgtgaag 4140
accttgtcag agggaataaa gcgaaagctg tgctttgtgc tgagcatcct ggggaacccg 4200
tcagtggtgc ttctggatga gccgtcgacc gggatggacc ccgaggggca gcagcaaatg 4260
tggcaggtga ttcgggccac ctttagaaac acggagaggg gcgccctcct gaccacccac 4320
tacatggcag aggctgaggc ggtgtgtgac cgagtggcca tcatggtgtc aggaaggctg 4380
agatgtattg gttccatcca acacctgaaa agcaaatttg gcaaagacta cctgctggag 4440
atgaagctga agaacctggc acaaatggag cccctccatg cagagatcct gaggcttttc 4500
ccccaggctg ctcagcagga aaggttctcc tccctgatgg tctataagtt gcctgttgag 4560
gatgtgcgac ctttatcaca ggctttcttc aaattagaga tagttaaaca gagtttcgac 4620
ctggaggagt acagcctctc acagtctacc ctggagcagg ttttcctgga gctctccaag 4680
gagcaggagc tgggtgatct tgaagaggac tttgatccct cggtgaagtg gaaactcctc 4740
ctgcaggaag agccttaaag ctccaatacc ctatatcttt ctttaatcct gtgactcttt 4800
taaagataat attttatagc cttaatatgc cttatatcag aggtggtaca aatgcatttg 4860
aaactcatgc caataattat cctcagtagt atttcttaca gtgagacaca ggcgatgtca 4920
gtgagggcga tcatagggca taagcctaag ccataccatg cagcctttgt gccagcaacc 4980
aatcccatgt ttcctactgt gttaagttta aaaatgcatt tattatagaa ttgtctacat 5040
ttctgaggat gtcatggaga atgcttaatt ttcttcctct gaacttcaaa atattaaata 5100
ttttcttatt tttttgatta aagtataaat taagacaccc tattgactt             5149
```